

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 BEVVPXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141691 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	69.2	426	1 AROA_VIBPA	Q87qx9 vibrio para
2	36	69.2	1057	1 CARB_STAPP	Q8cpj4 staphylococ
3	35	67.3	1058	1 CARB_FUSNN	Q8rg86 fusobacteri
4	35	67.3	1396	1 VITF_BPT5	P13390 bacterioph
5	34	65.4	102	1 Y1LK_TYDVA	P31619 tobacco yel
6	34	65.4	126	1 TRNK_BOVIN	P08858 bos taurus
7	34	65.4	1057	1 CARB_STAAM	Q99ur5 staphylococ
8	34	65.4	1057	1 CARB_STAAM	P58940 staphylococ
9	33	63.5	124	1 REV_STVCZ	P17280 chimpanzee
10	33	63.5	460	1 UMEI_YEAST	Q03010 saccharomyc
11	33	63.5	578	1 MDLB_BUCBP	Q89a96 buchnera ap
12	33	63.5	743	1 YK47_YEAST	P36148 saccharomyc
13	32	61.5	212	1 MSRA_VIBPA	Q87sw6 vibrio para
14	32	61.5	308	1 GAAB_METHF	Q26806 methanobact
15	32	61.5	363	1 ALFB_SHEEP	P52210 ovis aries
16	32	61.5	573	1 SUOX_DRONE	Q9vwp4 drosophila
17	32	61.5	627	1 MUTL_BACSU	P49850 bacillus su
18	32	61.5	735	1 ETRI_BRAOL	O49230 brassica ol
19	32	61.5	738	1 ETRI_AEATH	P49333 arabidopsis
20	32	61.5	906	1 CNC_MOUSE	P49452 mus musculu
21	32	61.5	1374	1 YS6_CAEEL	Q9541 caenorhabdi
22	32	61.5	1378	1 RON_MOUSE	Q62190 mus musculu
23	32	61.5	1454	1 PFTT_MOUSE	Q99m80 mus musculu
24	32	61.5	1463	1 PFTT_HUMAN	O14522 homo sapien
25	31	59.6	130	1 SZ05_RAT	P97885 rattus norv
26	31	59.6	212	1 MSRA_VIBCH	Q9kp30 vibrio chol
27	31	59.6	264	1 DPBB_WEITH	O27902 methanobact
28	31	59.6	319	1 YPAI_CRIPA	P10941 cryphonectr
29	31	59.6	363	1 ALFB_HUMAN	P05062 homo sapien
30	31	59.6	363	1 ALFB_RABIT	P79226 oryctolagus
31	31	59.6	366	1 DDL_OCEIH	Q8er16 oceanobacil
32	31	59.6	387	1 THIK_ECO57	Q8x814 escherichia
33	31	59.6	387	1 THIK_ECOLI	P21151 escherichia

34	31	59.6	387	1 THIK_SALTY	Q91616 salmonella
35	31	59.6	426	1 SLS1_YARLI	Q99158 yarrowia li
36	31	59.6	427	1 TOLB_HABIN	P44677 haemophilus
37	31	59.6	441	1 YLIS_VIBPA	P46231 vibrio para
38	31	59.6	991	1 DHP1_SCHPO	P40848 schizosacch
39	31	59.6	1188	1 KPBA_CAEEL	P34335 caenorhabdi
40	31	59.6	1410	1 PDRB_YEAST	P40550 saccharomyc
41	31	59.6	1499	1 A10A_HUMAN	O60312 homo sapien
42	31	59.6	3099	1 POLG_PEMVM	O56075 p genome po
43	30.5	58.7	268	1 ILIA_BOVIN	P08831 bos taurus
44	30.5	58.7	268	1 ILIA_CAPHI	P79161 capra hircu
45	30.5	58.7	268	1 ILIA_SHEEP	Q28579 ovis aries

ALIGNMENTS

RESULT 1
AROA_VIBPA
ID AROA_VIBPA STANDARD; PRT; 426 AA.
AC Q87QX9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR VF1020.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
sixth step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the EPSP synthase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AP005076; BAC59283.1; --
DR HAMAP; MF_00210; -- 1.
DR InterPro; IPR001986; EPSP synth.
DR Pfam; PF00275; EPSP synthase; 1.
DR PROSITE; PS00104; EPSP SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE_2; 1.
SQ Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
KW SEQUENCE 426 AA; 46094 MW; 373D39CC5BA1F70F CRC64;

Query Match 69.2%; Score 36; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 5.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGXHY 10
| | | | |
DB 223 EFVIPAGQHY 232

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 EVVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%
Maximum Watch 100%
Listing first 45 summaries

Database : A_Geneseq_25Jan04: *
1: Geneseq1980s: *
2: Geneseq1990s: *
3: Geneseq2000s: *
4: Geneseq2001s: *
5: Geneseq2002s: *
6: Geneseq2003as: *
7: Geneseq2003bs: *
8: Geneseq2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	5	ABB80525 Hepatitis
2	50	96.2	11	5	ABB80521 Hepatitis
3	50	96.2	11	5	ABB80522 Hepatitis
4	50	96.2	11	5	ABB80566 Hepatitis
5	50	96.2	11	5	ABB80563 Hepatitis
6	50	96.2	11	5	ABB80565 Hepatitis
7	50	96.2	11	5	ABB80567 Hepatitis
8	50	96.2	11	5	ABB80559 Hepatitis
9	50	96.2	11	5	ABB80526 Hepatitis
10	50	96.2	11	5	ABB80526 Hepatitis
11	50	96.2	11	5	ABB80568 Hepatitis
12	46	88.5	11	5	ABB80561 Hepatitis
13	46	88.5	11	5	ABB80524 Hepatitis
14	46	88.5	11	5	ABB80529 Hepatitis
15	46	88.5	11	5	ABB80528 Hepatitis
16	46	88.5	11	5	ABB80528 Hepatitis
17	45	86.5	11	5	ABB80523 Hepatitis
18	45	86.5	11	5	ABB80536 Hepatitis
19	45	86.5	11	5	ABB80558 Hepatitis
20	45	86.5	11	5	ABB80560 Hepatitis
21	45	86.5	11	5	ABB80527 Hepatitis
22	45	86.5	11	5	ABB80535 Hepatitis
23	45	86.5	11	5	ABB80540 Hepatitis
24	45	86.5	11	5	ABB80539 Hepatitis
25	44	84.6	11	5	ABB80549 Hepatitis

26	44	84.6	11	5	ABB80544 Hepatitis
27	44	84.6	11	5	ABB80553 Hepatitis
28	44	84.6	11	5	ABB80552 Hepatitis
29	44	84.6	11	5	ABB80545 Hepatitis
30	42	80.8	11	5	ABB80530 Hepatitis
31	41	78.8	11	5	ABB80542 Hepatitis
32	41	78.8	11	5	ABB80543 Hepatitis
33	41	78.8	11	5	ABB80538 Hepatitis
34	40	76.9	11	5	ABB80548 Hepatitis
35	40	76.9	11	5	ABB80547 Hepatitis
36	40	76.9	11	5	ABB80556 Hepatitis
37	40	76.9	11	5	ABB80557 Hepatitis
38	40	76.9	11	5	ABB80537 Hepatitis
39	40	76.9	11	5	ABB80551 Hepatitis
40	40	76.9	11	5	ABB80541 Hepatitis
41	40	76.9	20	2	AAU78810 Hepatitis
42	40	76.9	1022	4	ABG03621 Novel hum
43	40	76.9	1022	4	ABG08173 Novel hum
44	40	76.9	1022	4	ABG05826 Novel hum
45	39	75.0	11	5	ABB80546 Hepatitis

ALIGNMENTS

RESULT 1

ID	ABB80525	standard; peptide; 11 AA.
XX	ABB80525;	
AC	ABB80525;	
XX		
DT	08-OCT-2002	(first entry)
XX		
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.	
XX		
DE	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.	
KM		
KW		
XX		
OS	Synthetic.	
XX		
EH	Key	Location/Qualifiers
FT	Modified-site 1	/note= "N-terminal acetyl"
FT	Modified-site 6	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 8	/note= "D-form residue"
FT	Modified-site 11	/note= "C-terminal amide"
XX		
XX	WO200208251-A2.	
XX	31-JAN-2002.	
XX	19-JUL-2001; 2001WO-US023169.	
XX	21-JUL-2000; 2000US-0220101P.	
PR	(CORV-) CORVAS INT INC.	
PA	Lim-Wilby M, Levy OE, Brunck TK;	
PI	WPI; 2002-361643/39.	
DR		
XX		
XX		
PT		
PT		
XX		
XX		
PS		
XX		
CC		
CC		

Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 2
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.

AC ABB80521;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX Claim 17; Page 64; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 3
 ABB80522
 ID ABB80522 standard; peptide; 11 AA.

AC ABB80522;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 4

BB80566
D ABB80566 standard; peptide; 11 AA.
X C
X C
X C
T 08-OCT-2002 (first entry)
X X
E Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
X X
W Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
X virucide.
X X
S Synthetic.
X X
H Key
T Modified-site 1 Location/Qualifiers
T /note= "N-terminal acetyl"
T Modified-site 6
T /note= "2-aminoisobutyl carbonyl residue forming a keto
T -amide linkage with residue 7"
T Modified-site 11
T /note= "C-terminal amide"
T X
N WO200208251-A2.
X N
X 31-JAN-2002.
X D
X 19-JUL-2001; 2001WO-US023169.
X F
X 21-JUL-2000; 2000US-0220101P.
X R
X (CORV-) CORVAS INT INC.
X A
X Lim-Wilby M, Levy OE, Brunck TK;
X WPI; 2002-361643/39.
X X
X Novel peptide compound having hepatitis C virus protease inhibitory
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X T
X Claim 17; Page 65; 69pp; English.
X S
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus
X
X Sequence 11 AA;
X
X Query Match 96.2%; Score 50; DB 5; Length 11;
X Best Local Similarity 100.0%; Pred. No. 0.002;
X Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
X
X Y 1 EEVVPXGMSYS 11
X | | | | | | | | | |
X 1 EEVVPXGMSYS 11
X
X Db
X
X RESULT 5
X ABB80563
X ID ABB80563 standard; peptide; 11 AA.
X X
X ABB80563;
X X
X 08-OCT-2002 (first entry)
X X
X Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.
X X
X Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
X X

virucide.
XW
X Synthetic.
X X
X Key
T Modified-site 1 Location/Qualifiers
T /note= "N-terminal acetyl"
T Modified-site 6
T /note= "valyl carbonyl forming keto-amide linkage with
T residue 7"
T Modified-site 11
T /note= "C-terminal amide"
T X
X WO200208251-A2.
X P
X 31-JAN-2002.
X X
X 19-JUL-2001; 2001WO-US023169.
X PF
X 21-JUL-2000; 2000US-0220101P.
X PR
X (CORV-) CORVAS INT INC.
X PA
X Lim-Wilby M, Levy OE, Brunck TK;
X WPI; 2002-361643/39.
X DR
X Novel peptide compound having hepatitis C virus protease inhibitory
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X PT
X Claim 17; Page 65; 69pp; English.
X PS
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus
X
X Sequence 11 AA;
X
X Query Match 96.2%; Score 50; DB 5; Length 11;
X Best Local Similarity 100.0%; Pred. No. 0.002;
X Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
X
X Qy 1 EEVVPXGMSYS 11
X | | | | | | | | | |
X 1 EEVVPXGMSYS 11
X
X Db
X
X RESULT 6
X ABB80565
X ID ABB80565 standard; peptide; 11 AA.
X X
X ABB80565;
X X
X 08-OCT-2002 (first entry)
X DT
X X
X DE
X X
X Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
X X
X Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
X virucide.
X X
X Synthetic.
X OS
X Key
T Modified-site 1 Location/Qualifiers
T /note= "N-terminal acetyl"
T Modified-site 6
T /note= "Norleucyl carbonyl forming keto-amide linkage
T with residue 7"
T FT

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 CX
 PS Claim 17; Page 65; 69pp; English.
 CX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 CX
 SQ Sequence 11 AA;
 Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 DB 1 BEVVPXGMSYS 11
 RESULT 9
 ID ABB80526 standard; peptide; 11 AA.
 CX ABB80526;
 CX 08-OCT-2002 (first entry)
 CX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
 CX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CX virucide.
 CX Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 CX
 CX 31-JAN-2002.
 CX 19-JUL-2001; 2001WO-US023169.
 CX 21-JUL-2000; 2000US-0220101P.
 CX (CORV-) CORVAS INT INC.
 CX Lim-Wilby M, Levy OE, Brunck TK;
 CX WPI; 2002-361643/39.
 CX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 CX
 PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 CX
 SQ Sequence 11 AA;
 Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 DB 1 BEVVPXGMSYS 11
 RESULT 10
 ID ABB80564 standard; peptide; 11 AA.
 CX ABB80564;
 CX 08-OCT-2002 (first entry)
 CX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
 CX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CX virucide.
 CX Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 CX
 CX 31-JAN-2002.
 CX 19-JUL-2001; 2001WO-US023169.
 CX 21-JUL-2000; 2000US-0220101P.
 CX (CORV-) CORVAS INT INC.
 CX Lim-Wilby M, Levy OE, Brunck TK;
 CX WPI; 2002-361643/39.
 CX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 CX
 PS Claim 17; Page 65; 69pp; English.
 CX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 CX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 Db 1 BEVVPXGMSYS 11

RESULT 11
 ABB80568
 ID ABB80568 standard; peptide; 11 AA.
 XX AC ABB80568;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.
 XX FH Key
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
 FT FT a keto-amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT FT
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
 Claim 17; Page 65; 69pp; English.
 The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 Db 1 BEVVPXGMSYS 11

RESULT 12
 ABB80561
 ID ABB80561 standard; peptide; 11 AA.
 XX AC ABB80561;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.

XX FH Key
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT FT
 XX WO200208251-A2.

XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 Db 1 BEVVPXGMSYS 11

RESULT 13
 ABB80524
 ID ABB80524 standard; peptide; 11 AA.
 XX AC ABB80524;
 XX XX

08-OCT-2002 (first entry)
Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
Synthetic.

Key Location/Qualifiers
Modified-site 1 /note= "N-terminal acetyl"
Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
Misc-difference 8 /note= "D-form residue"
Misc-difference 9 /note= "D-form residue"
Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
Claim 17; Page 64; 69pp; English.
The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
| | | | | | | |
Db 1 EEVVPXGMDYS 11
| | | | | | | |

RESULT 14
ABB80529
ID ABB80529 standard; peptide; 11 AA.
XX ABB80529;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
XX
OS Synthetic.

XX Key Location/Qualifiers
PH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
Claim 17; Page 64; 69pp; English.
The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
| | | | | | | |
Db 1 EEVVPXGMDYS 11
| | | | | | | |

RESULT 15
ABB80528
ID ABB80528 standard; peptide; 11 AA.
XX ABB80528;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
XX
OS Synthetic.

Key Location/Qualifiers
PH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8 residue 7"
FT /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
XX
DN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ

Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
| | | | | | | | | |
Db 1 BEVVPXGNDYS 11

Search completed: June 3, 2004, 11:48:22
Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-5

Perfect score: 52

Sequence: 1 EVVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents AA:*
- 2: /cgn2_6/ptodata/2/iaa/5A COMB.pcp:*
- 3: /cgn2_6/ptodata/2/iaa/5B COMB.pcp:*
- 4: /cgn2_6/ptodata/2/iaa/6A COMB.pcp:*
- 5: /cgn2_6/ptodata/2/iaa/6B COMB.pcp:*
- 6: /cgn2_6/ptodata/2/iaa/PTCUT COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	4	US-09-408-020-4
2	36	69.2	382	4	US-09-134-000C-3738
3	34	65.4	947	4	US-09-228-986-73
4	34	65.4	1191	4	US-09-540-236-2902
5	34	65.4	1407	4	US-09-328-352-7885
6	33	63.5	12	4	US-09-760-946-2
7	33	63.5	12	4	US-09-760-946-3
8	33	63.5	45	2	US-08-637-759B-236
9	33	63.5	45	3	US-08-871-355A-236
10	33	63.5	45	4	US-09-201-945-236
11	33	63.5	65	6	517197-51
12	33	63.5	410	6	517197-1
13	33	63.5	1394	6	517197-30
14	32	61.5	10	3	US-09-357-952-66
15	32	61.5	10	4	US-09-521-650-66
16	32	61.5	10	4	US-09-168-888-66
17	32	61.5	102	2	US-08-580-988A-23
18	32	61.5	152	2	US-08-460-694-4
19	32	61.5	152	3	US-08-460-744-4
20	32	61.5	152	3	US-07-667-711B-4
21	32	61.5	173	1	US-08-193-377-7
22	32	61.5	189	2	US-08-464-517-21
23	32	61.5	189	2	US-08-246-361A-21
24	32	61.5	189	3	US-08-463-772-21
25	32	61.5	189	5	PCT-US93-05000-21
26	32	61.5	236	2	US-08-454-517-22
27	32	61.5	236	2	US-08-246-361A-22

28 32 61.5 236 3 US-08-463-772-22 Sequence 22, Appl
29 32 61.5 236 5 PCT-US93-05000-22 Sequence 22, Appl
30 32 61.5 280 2 US-08-464-517-6 Sequence 6, Appl
31 32 61.5 280 3 US-08-463-772-6 Sequence 6, Appl
32 32 61.5 289 2 US-08-246-361A-4 Sequence 4, Appl
33 32 61.5 289 5 PCT-US93-05000-4 Sequence 4, Appl
34 32 61.5 291 5 PCT-US93-05000-6 Sequence 6, Appl
35 32 61.5 292 2 US-08-464-517-23 Sequence 23, Appl
36 32 61.5 292 2 US-08-246-361A-6 Sequence 23, Appl
37 32 61.5 292 2 US-08-246-361A-23 Sequence 23, Appl
38 32 61.5 292 3 US-08-463-772-23 Sequence 23, Appl
39 32 61.5 292 5 PCT-US93-05000-23 Sequence 23, Appl
40 32 61.5 295 1 US-07-947-120-8 Sequence 8, Appl
41 32 61.5 295 1 US-08-472-893A-8 Sequence 2, Appl
42 32 61.5 295 2 US-08-460-694-2 Sequence 19, Appl
43 32 61.5 295 2 US-08-464-517-19 Sequence 20, Appl
44 32 61.5 295 2 US-08-464-517-20 Sequence 19, Appl
45 32 61.5 295 2 US-08-246-361A-19 Sequence 19, Appl

ALIGNMENTS

RESULT 1

US-09-408-020-4

; Sequence 4, Application US/09408020

; Patent No. 6632937

; GENERAL INFORMATION:

; APPLICANT: Swanson, Ronald V.

; APPLICANT: Feldman, Robert A.

; APPLICANT: Schleper, Christa

; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM

; FILE REFERENCE: DCOEP 002A

; CURRENT APPLICATION NUMBER: US/09/408,020

; PRIOR FILING DATE: 1999-09-29

; PRIOR FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 123

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 3472

; TYPE: PRT

; ORGANISM: Cenarchaeum symbiosum

US-09-408-020-4

Query Match 73.1%; Score 38; DB 4; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11

Db 2294 EDVPRGIGSFS 2304

RESULT 2

US-09-134-000C-3738

; Sequence 3738, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 012796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3738

; LENGTH: 382

; TYPE: PRT

; ORGANISM: Enterococcus faecalis


```
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (327)..(328)
/ OTHER INFORMATION: Amino acids 327 & 328 are Xaa wherein Xaa = any amino acid.
US-09-134-000C-3738

Query Match      69.2%; Score 36; DB 4; Length 382;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVVPXGMSYS 11
      :|||
Db      332 LIPEGMSYS 340

RESULT 3
US-09-228-986-73
/ Sequence 73, Application US/09228986
/ Patent No. 6359198
/ GENERAL INFORMATION:
/ APPLICANT: Strabala, Timothy
/ APPLICANT: Nieuwenhuizen, Niels
/ TITLE OF INVENTION: Compositions Isolated from Plant Cells
/ FILE REFERENCE: 11000/1020
/ CURRENT APPLICATION NUMBER: US/09/228,986
/ CURRENT FILING DATE: 1999-01-12
/ NUMBER OF SEQ ID NOS: 130
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 73
/ LENGTH: 947
/ TYPE: PRT
/ ORGANISM: Pinus radiata
US-09-228-986-73

Query Match      65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVVPXGMSYS 11
      :|||
Db      686 VMPGMSYS 694

RESULT 4
US-09-540-236-2902
/ Sequence 2902, Application US/09540236
/ Patent No. 6673910
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
/ FILE REFERENCE: 2709.2005-001
/ CURRENT APPLICATION NUMBER: US/09/540,236
/ CURRENT FILING DATE: 2000-04-04
/ NUMBER OF SEQ ID NOS: 3840
/ SEQ ID NO 2902
/ LENGTH: 1191
/ TYPE: PRT
/ ORGANISM: M. catarrhalis
US-09-540-236-2902

Query Match      65.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      2 EVVPXGMSY 10
      :|||
Db      783 EILPVGMAY 791

RESULT 5
US-09-328-352-7885
/ Sequence 7885, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ FILE REFERENCE: CTC99-03EA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 7885
/ LENGTH: 1407
/ TYPE: PRT
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match      65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 EVVPXGMSY 10
      :|||
Db      596 EVPEGLSP 604

RESULT 6
US-09-760-946-2
/ Sequence 2, Application US/09760946
/ Patent No. 6608027
/ GENERAL INFORMATION:
/ APPLICANT: Tsantrizos, Youla S.
/ APPLICANT: Cameron, Dale R.
/ APPLICANT: Faucher, Anne-Marie
/ APPLICANT: Ghio, Elise
/ APPLICANT: Goudreau, Nathalie
/ APPLICANT: Halmos, Teddy
/ APPLICANT: Llinas-Brunet, Montse
/ TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
/ FILE REFERENCE: 13/076-1-C1
/ CURRENT APPLICATION NUMBER: US/09/760,946
/ CURRENT FILING DATE: 2001-08-23
/ PRIOR APPLICATION NUMBER: US 09/542,675
/ PRIOR FILING DATE: 2000-04-03
/ PRIOR APPLICATION NUMBER: US 60/128,011
/ PRIOR FILING DATE: 1999-04-06
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 2
/ LENGTH: 12
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Substrate for recombinant HCV NS3 protease radiometric assay
US-09-760-946-2

Query Match      63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
      :|||
Db      1 DDIVPCMSYS 11

RESULT 7
US-09-760-946-3
/ Sequence 3, Application US/09760946
/ Patent No. 6608027
/ GENERAL INFORMATION:
/ APPLICANT: Tsantrizos, Youla S.
/ APPLICANT: Cameron, Dale R.
/ APPLICANT: Faucher, Anne-Marie
/ APPLICANT: Ghio, Elise
```

```

; APPLICANT: Goudreau, Nathalie
; APPLICANT: Halmos, Teddy
; APPLICANT: Llinas-Bruner, Montse
; TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
; FILE REFERENCE: 13/076-1-C1
; CURRENT APPLICATION NUMBER: US/09/760,946
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/542,675
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/128,011
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tracer for NS3 protease assay
; NAME/KEY: MCD RES
; LOCATION: (1)
; OTHER INFORMATION: Asp at position 1 is biotinylated
; NAME/KEY: MCD RES
; LOCATION: (10)
; OTHER INFORMATION: Tyr at position 10 is iodinated with I-125
JS-09-760-946-3

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

2Y 1 BEVVPXGMSYS 11
   |||
   |||
DB 1 DDIVPCMSYT 11

```

```

RESULT 8
JS-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
JS-08-637-759B-236

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
   |||
   |||
DB 1 EEISPLGWSY 10

```

```

RESULT 9
US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
   |||
   |||
DB 1 EEISPLGWSY 10

```

RESULT 10
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
Db 1 BEISPLGWSY 10

RESULT 11
5177197-51
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAES-
SON-WELSH, LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 51:
; LENGTH: 65

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 12;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
Db 52 KEICPGMGYT 62

RESULT 12
5177197-1
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAES-
SON-WELSH, LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 1:
; LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 97;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
Db 399 KEICPGMGYT 409

RESULT 13
5177197-30
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAES-
SON-WELSH, LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 30:
; LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
Db 399 KEICPGMGYT 409

RESULT 14
US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whole-
; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Protease;
; TITLE OF INVENTION: Other Enzymes and the Use Thereof
; FILE REFERENCE: 1735.0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642

EARLIER FILING DATE: 21-JUL-1998
 NUMBER OF SEQ ID NOS: 139
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 66
 LENGTH: 10
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 OTHER INFORMATION: Peptide
 S-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

y 1 EVVVPXGMSY 10
 : : : : :
 b 1 DDIVPCMSY 10

RESULT 15
 S-09-521-650-66
 Sequence 66, Application US/09521650
 Patent No. 6335429
 GENERAL INFORMATION:
 APPLICANT: Weber Eckard
 APPLICANT: Cai, Sui Xiong
 APPLICANT: Keana, John F.W.
 APPLICANT: Drewe, John A.
 APPLICANT: Zhang, Han-Zhong
 TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
 TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
 TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
 TITLE OF INVENTION: Use Thereof
 FILE REFERENCE: 1735.0290002
 CURRENT APPLICATION NUMBER: US/09/521,650
 CURRENT FILING DATE: 2000-03-08
 EARLIER APPLICATION NUMBER: 09/168,888
 EARLIER FILING DATE: 1998-10-09
 EARLIER APPLICATION NUMBER: US 60/061,582
 EARLIER FILING DATE: 1997-10-10
 EARLIER APPLICATION NUMBER: US 09/033,661
 EARLIER FILING DATE: 1998-03-03
 NUMBER OF SEQ ID NOS: 142
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 66
 LENGTH: 10
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 OTHER INFORMATION: Peptide
 S-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

y 1 EVVVPXGMSY 10
 : : : : :
 b 1 DDIVPCMSY 10

Search completed: June 3, 2004, 12:03:06
 Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1155919 seqs, 281338677 residues
Total number of hits satisfying chosen parameters: 1155919
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	96.2	11	12	US-09-909-164-5
2	50	96.2	11	12	US-09-909-164-6
3	50	96.2	11	12	US-09-909-164-9
4	50	96.2	11	12	US-09-909-164-10
5	50	96.2	11	12	US-09-909-164-47
6	50	96.2	11	12	US-09-909-164-48
7	50	96.2	11	12	US-09-909-164-49
8	50	96.2	11	12	US-09-909-164-50
9	50	96.2	11	12	US-09-909-164-51
10	50	96.2	11	12	US-09-909-164-52
11	46	88.5	11	12	US-09-909-164-8
12	46	88.5	11	12	US-09-909-164-12
13	46	88.5	11	12	US-09-909-164-13
14	45	86.5	11	12	US-09-909-164-7
15	45	86.5	11	12	US-09-909-164-11

16	45	86.5	11	12	US-09-909-164-19
17	45	86.5	11	12	US-09-909-164-20
18	45	86.5	11	12	US-09-909-164-23
19	45	86.5	11	12	US-09-909-164-24
20	44	84.6	11	12	US-09-909-164-28
21	44	84.6	11	12	US-09-909-164-29
22	44	84.6	11	12	US-09-909-164-33
23	44	84.6	11	12	US-09-909-164-36
24	44	84.6	11	12	US-09-909-164-37
25	44	84.6	11	12	US-09-909-164-43
26	42	80.8	11	12	US-09-909-164-14
27	41	78.8	11	12	US-09-909-164-22
28	41	78.8	11	12	US-09-909-164-26
29	41	78.8	11	12	US-09-909-164-27
30	41	78.8	11	12	US-09-909-164-61
31	41	78.8	11	12	US-09-909-164-62
32	40	76.9	11	12	US-09-909-164-21
33	40	76.9	11	12	US-09-909-164-25
34	40	76.9	11	12	US-09-909-164-31
35	40	76.9	11	12	US-09-909-164-32
36	40	76.9	11	12	US-09-909-164-35
37	40	76.9	11	12	US-09-909-164-40
38	40	76.9	11	12	US-09-909-164-41
39	40	76.9	11	12	US-09-909-164-45
40	40	76.9	11	12	US-09-909-164-46
41	39	75.0	11	12	US-09-909-164-30
42	39	75.0	11	12	US-09-909-164-34
43	39	75.0	11	12	US-09-909-164-38
44	39	75.0	11	12	US-09-909-164-39
45	39	75.0	11	12	US-09-909-164-42

ALIGNMENTS

RESULT 1
US-09-909-164-5
; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-5
Query Match 96.2% ; Score 50 ; DB 12 ; Length 11 ;

```
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 2
US-09-909-164-6
; Sequence 6, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-6
; Query Match 96.2%; Score 50; DB 12; Length 11;
; Best Local Similarity 100.0%; Pred. No. 0.0014;
; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 3
US-09-909-164-9
; Sequence 9, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
```

```
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-9
; Query Match 96.2%; Score 50; DB 12; Length 11;
; Best Local Similarity 100.0%; Pred. No. 0.0014;
; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 4
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
```

OTHER INFORMATION: D-amino acids
US-09-909-164-10

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
|||||
DB 1 BEVVPXGMSYS 11

RESULT 5

US-09-909-164-47

Sequence 47, Application US/09909164
Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunn, Terence K

FILE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

TITLE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

PRIOR FILING DATE: 2003-03-25

PRIOR FILING DATE: 2003-03-25

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 47

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLATION

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: valine-(CO)

US-09-909-164-47

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
|||||
DB 1 BEVVPXGMSYS 11

RESULT 6

US-09-909-164-48

Sequence 48, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunn, Terence K

FILE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

TITLE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

PRIOR FILING DATE: 2003-03-25

PRIOR FILING DATE: 2003-03-25

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: leucine-(CO)
US-09-909-164-48

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
|||||
DB 1 BEVVPXGMSYS 11

RESULT 7

US-09-909-164-49

Sequence 49, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunn, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

CURRENT FILING DATE: 2003-03-25

PRIOR APPLICATION NUMBER: 60/220,101

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 49

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLATION

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norleucine-(CO)

US-09-909-164-49

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11

```
Db      1  EVVVPXGMSYS 11
|||||
RESULT 8
US-09-909-164-50
; Sequence 50, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: (s,s)-allothreonline-(CO)
US-09-909-164-51
Query Match          96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  EVVVPXGMSYS 11
|||||
Db      1  EVVVPXGMSYS 11
|||||
RESULT 10
US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: propargyl glycine-(CO)
US-09-909-164-53
Query Match          96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  EVVVPXGMSYS 11
|||||
Db      1  EVVVPXGMSYS 11
|||||
RESULT 11
US-09-909-164-54
; Sequence 54, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: propargyl glycine-(CO)
US-09-909-164-55
Query Match          96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  EVVVPXGMSYS 11
|||||
Db      1  EVVVPXGMSYS 11
|||||
RESULT 9
US-09-909-164-51
; Sequence 51, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
```



```
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-8

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 EVVVPXGMSYS 11
  |||||
b 1 EVVVPXGMDYS 11

RESULT 12
US-09-909-164-12
Sequence 12, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-13

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 EVVVPXGMSYS 11
  |||||
b 1 EVVVPXGMDYS 11

RESULT 13
US-09-909-164-13
Sequence 13, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-13

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 EVVVPXGMSYS 11
  |||||
b 1 EVVVPXGMDYS 11
```

RESULT 14

```
US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication NO. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-7
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 EEVVPXGMSYS 11
   |||||
Db 1 EEVVPXGMHYS 11
```

RESULT 15

```
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication NO. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
```

```
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-11
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 EEVVPXGMSYS 11
   |||||
Db 1 EEVVPXGMHYS 11
```

```
Search completed: June 3, 2004, 12:57:14
Job time : 33.7333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

MM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-5

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 781*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2 T31308	hypothetical 367K
2	37	71.2	840	2 T39116	probable sulfate p
3	37	71.2	877	2 T40413	sulfate permease -
4	36	69.2	102	2 A42452	V1 protein - tobac
5	36	69.2	1498	2 B97355	DNA segregation AT
6	35	67.3	225	2 S57810	hypothetical prote
7	35	67.3	425	2 T24111	hypothetical prote
8	35	67.3	670	2 S22293	zinc finger protei
9	35	67.3	749	2 H82691	topoisomerase IV s
10	35	67.3	2717	2 A34203	DNA-binding protei
11	34	65.4	156	2 S54619	hypothetical prote
12	34	65.4	252	2 H69491	cell division inhi
13	34	65.4	544	2 C82900	probable ABC subu
14	33	63.5	94	2 T40758	hypothetical prote
15	33	63.5	116	2 B90544	50S ribosomal prot
16	33	63.5	165	2 D69493	hypothetical prote
17	33	63.5	253	2 C81374	hypothetical prote
18	33	63.5	259	2 T34536	hypothetical prote
19	33	63.5	284	2 S75817	hypothetical prote
20	33	63.5	298	2 T47670	beta-ketocacyl-ACP
21	33	63.5	368	2 F72281	hypothetical prote
22	33	63.5	426	2 D82163	3-phosphoshikimate
23	33	63.5	466	2 T43653	cdc37 protein - fi
24	33	63.5	653	2 D82352	iron(III) ABC tran
25	33	63.5	890	2 A30481	bacteriocin BGN5 -
26	33	63.5	1028	2 AP3286	ATP-dependent DNA
27	33	63.5	1152	2 D87046	conserved hypothet
28	33	63.5	1394	2 A35626	transforming growt
29	33	63.5	1401	2 G82336	DNA-directed RNA p

30	33	63.5	1548	2 T04456	hypothetical prote
31	33	63.5	1712	2 A38261	masking protein pr
32	32	61.5	84	2 E97333	hypothetical prote
33	32	61.5	175	2 PQ0616	transport protein
34	32	61.5	223	2 T01457	rho protein GDP-di
35	32	61.5	279	2 B72481	hypothetical prote
36	32	61.5	288	2 JC0411	cyclin D2 - rat
37	32	61.5	288	2 I59372	cyclin D2 - mouse
38	32	61.5	289	2 A41984	cyclin D2 - human
39	32	61.5	289	2 A42822	cyclin D2 - Africa
40	32	61.5	291	2 S57922	cyclin D2 - chicke
41	32	61.5	291	2 S57925	cyclin D2 - zebra
42	32	61.5	291	2 JC4579	cyclin D1 - human
43	32	61.5	291	2 S62730	cyclin D3 - human
44	32	61.5	292	2 B42822	cyclin D1 - human
45	32	61.5	295	2 A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C;Species: Cenarchaeum symbiosum
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C;Accession: T31308
J;Schleper, C.; DeLong, E.P.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J;Bacteriol. 180, 5003-5009, 1998
A;Title: Genomic analysis reveals chromosomal variation in natural populations of the ur
A;Reference number: Z20994; MUID:98422450; PMID:9748430
A;Accession: T31308
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3472 <SCH>
A;Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599394; PIDN:AAC62699.1
C;Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;

Best Local Similarity 54.5%; Pred. No. 60;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11

Db 2294 EDVIPRGISFS 2304

RESULT 2

T39116
Probable sulfate permease - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39116
R;Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A;Reference number: Z21829
A;Accession: T39116
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-840 <HUN>
A;Cross-references: EMBL:AL132779; PIDN:CAB60015.1; GSPDB:GN00066; SPDB:SPAC869.05C
A;Experimental source: strain 972h; cosmid c869
C;Genetics:
A;Gene: SPDB:SPAC869.05C
A;Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;

Best Local Similarity 77.8%; Pred. No. 21;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11

Db 135 VVPGMSYA 143

RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40413
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z21926
A;Accession: T40413
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-877 <LYN>
A;Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A;Experimental source: strain 972h-; cosmid c3H7
C;Genetics:
A;Gene: SPDB:SPBC3H7.02
A;Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 148 VVPQGSYA 156
:|||||:

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C;Species: tobacco yellow dwarf virus
C;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C;Accession: A42452
R;Morris, B.A.M.; Richardson, K.A.; Hailey, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A;Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A;Reference number: A42452; MUID:92188538; PMID:1546458
A;Accession: A42452
A;Molecule type: DNA
A;Residues: 1-102 <MOR>
A;Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
DB 7 QVVPSGINS 16
:|||||:

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clo
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: B97355
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97355
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1498 <KUR>
A;Cross-references: GB:AE001437; PIDN:AK81629.1; PID:gl5026814; GSPDB:GN00169
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1276 EQIPMGMSY 1285
:|||||:

RESULT 6
S57810
hypothetical protein precursor (clone TPp11) - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C;Accession: S57810
R;Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A;Title: Nature and regulation of pistil-expressed genes in tomato.
A;Reference number: S57808; MUID:95375233; PMID:7647301
A;Accession: S57810
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-225 <MIL>
A;Cross-references: EMBL:U20592; NID:G924625; PIDN:AAA80497.1; PID:G924626
C;Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 32 DEVPVNGTYA 42
:|||||:

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111
R;Percy, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19842
A;Accession: T24111
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
C;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 335 EQIVFGGLQY 344
:|||||:

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 23-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C;Accession: S22293; I78656
R;Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

```

;Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
;Reference number: I58280; MUID:91187610; PMID:1901405
;Accession: S22933
;Status: nucleic acid sequence not shown
;Molecule type: mRNA
;Residues: 1-670 <MIT>
;Cross-references: EMBL:X54250; NID:G57519; PIDN:CAA38151.1; PID:G57520
;Note: The authors did not translate the codon for residue 1
;Superfamily: HIV-EP2 enhancer-binding protein
;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. NO. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Y 3 VVPXGMSYS 11
||| |||
b 376 VVPAGLTYS 384

RESULT 9
82691
opoisomerase IV subunit XF1353 [imported] - Xylella fastidiosa (strain 9aSc)
;Species: Xylella fastidiosa
;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
;Accession: H82691
;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
ature 406, 451-157, 2000
;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
;Reference number: A82515; MUID:20365717; PMID:10910347
;Note: for a complete list of authors see reference number A59328 below
;Accession: H82691
;Status: preliminary
;Molecule type: DNA
;Residues: 1-749 <SIM>
;Cross-references: GB:AE003967; GB:AE003849; NID:G9106347; PIDN:AAF84162.1; GSPDS:GN001
;Experimental source: strain 9aSc
;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-
riones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
s-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
hado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
odrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
;Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
l.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
;Reference number: A59328
;Contents: annotation
;Genetics:
;Gene: XF1353
;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phase T4 DNA topoisomerase
Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. NO. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 2 EVVPXGMSY 10
||| |||
b 526 EVDPFGMSY 534

RESULT 10
134203
NA-binding protein PRDII-BF1 - human
;Alternate names: major histocompatibility complex enhancer-binding protein 1
;Species: Homo sapiens (man)
;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
;Accession: A34203; A34779
;Fan, C.M.; Maniatis, T.
;Genes Dev. 4, 29-42, 1990

```

```

;Title: A DNA-binding protein containing two widely separated zinc finger motifs that r
;Reference number: A34203; MUID:90169514; PMID:2106471
;Accession: A34203
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-2717 <FAN>
;Cross-references: EMBL:X51435; NID:G38017; PIDN:CAA35798.1; PID:G38018
R.; Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
;Title: A large protein containing zinc finger domains binds to related sequence elemen
;Reference number: A34779; MUID:90205817; PMID:2108316
;Accession: A34779
;Status: preliminary; nucleic acid sequence not shown
;Molecule type: mRNA
;Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-16'
;Cross-references: GB:M32019
;Superfamily: HIV-EP2 enhancer-binding protein
;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. NO. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| |||
Db 2405 VVPAGLTYS 2413

RESULT 11
S54619
hypothetical protein YOR013W - yeast (Saccharomyces cerevisiae)
;Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
;Species: Saccharomyces cerevisiae
;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
;Accession: S54619; S66879
R.; de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
;Reference number: S54617
;Accession: S54619
;Molecule type: DNA
;Residues: 1-156 <DEH>
;Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123
R.; de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
;Reference number: S66877
;Accession: S66879
;Molecule type: DNA
;Residues: 1-156 <DEW>
;Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA99201.1; PID:G1420111; MIPS:YOR01.
;Experimental source: strain S288C
;Genetics:
;Cross-references: SGD:S0005539
;Map position: 15R
;Superfamily: hypothetical protein YOR013W

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. NO. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
||| |||
Db 50 EVNPLGMDY 58

RESULT 12
H69491
cell division inhibitor (minD-2) homolog - Archaeoglobus fulgidus
;Species: Archaeoglobus fulgidus
;Date: 03-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
;Accession: H69491
R.; Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

```

Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Wiese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AE000970; GB:AE000782; NID:92689293; PIDN:AA889318.1; PID:9264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMS 9
 |||||
 Db 81 EVIPAGMS 88
 |||||

RESULT 13
 C82900
 probable ABC substrate-binding protein, iron U0359 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
 A:Reference number: A82870
 A:Accession: C82900
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:96899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3, Biovar 1
 C:Genetics:
 A:Gene: ABCsbp-5; U0359
 A:Genetic code: GCG3

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEWVPXGMSY 10
 |||||
 Db 135 EEWVPHYLSY 144
 |||||

RESULT 14
 I40758
 hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995

A:Title: Expression and characterization of Campylobacter jejuni benzoyl-glycine amidohyd
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:236940; NID:9535805; PIDN:CAA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 |||||
 Db 26 DIFPSCMSY 34
 |||||

RESULT 15
 E90544
 50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Charbaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, P.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: E90544
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPU 2610
 A:Genetic code: GCG3
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 |||||
 Db 68 VRPLGMSYS 76
 |||||

Search completed: June 3, 2004, 11:59:58
 Job time : 10 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 EEVFXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	1058	1 CARB_FUSNN	Q8rg86 fusobacteri
2	37	71.2	877	1 SULH_SCHPO	C74377 schizosacch
3	36	69.2	102	1 Y11K_TYDVA	P31619 tobacco yel
4	36	69.2	1498	1 Y1A9_CLOAB	Q04351 clostridium
5	35	67.3	2717	1 ZEPI_HUMAN	P15822 homo sapien
6	34	65.4	788	1 CV14_NEUCR	P23622 neurospora
7	34	65.4	1499	1 A10A_HUMAN	O60312 homo sapien
8	33	63.5	116	1 RL20_MGCPU	Q98q70 mycoplasma
9	33	63.5	165	1 YJ49_ARCFU	C28330 archaescglob
10	33	63.5	253	1 Y990_CAMJE	P45489 campylobact
11	33	63.5	280	1 CTX3_MOUSE	Q9d387 mus musculu
12	33	63.5	426	1 AROA_VIBCH	Q9krb0 vibrio chol
13	33	63.5	466	1 CC37_SCHPO	O94740 schizosacch
14	33	63.5	478	1 GSR2_HUMAN	O9azm5 homo sapien
15	33	63.5	890	1 BCN1_CLOPE	P08696 clostridium
16	33	63.5	1389	1 LTBS_MOUSE	Q8eg18 mus musculu
17	33	63.5	1394	1 LTBS_MOUSE	P22064 homo sapien
18	33	63.5	1401	1 RPOC_VIBCH	Q9kv29 vibrio chol
19	33	63.5	1595	1 LTBL_HUMAN	Q14766 homo sapien
20	33	63.5	1712	1 LTBI_RAT	Q00918 rattus norv
21	33	63.5	1713	1 LTBL_MOUSE	Q8cg19 mus musculu
22	32	61.5	288	1 CGD2_RAT	Q04827 rattus norv
23	32	61.5	289	1 CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	1 CGD2_MOUSE	Q30280 mus musculu
25	32	61.5	291	1 CGD1_BRARE	Q90459 brachydanio
26	32	61.5	291	1 CGD1_XENLA	P50755 xenopus lae
27	32	61.5	291	1 CGD2_CHICK	P49706 gallus gall
28	32	61.5	291	1 CGD2_XENLA	P53782 xenopus lae
29	32	61.5	292	1 CGD1_CHICK	P55169 gallus gall
30	32	61.5	292	1 CGD3_HUMAN	P30281 homo sapien
31	32	61.5	295	1 CGD1_HUMAN	P24385 homo sapien
32	32	61.5	295	1 CGD1_MOUSE	P25322 mus musculu
33	32	61.5	295	1 CGD1_RAT	P39948 rattus norv

34	32	61.5	427	1 TOLB_HAEIN	P44677 haemophilus
35	32	61.5	529	1 ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1 S216_HUMAN	Q9v616 homo sapien
37	32	61.5	719	1 GSP_CRIFA	P90518 crithidia f
38	32	61.5	726	1 PRTP_HSV6U	P53384 human herpe
39	32	61.5	726	1 PRTP_HSV6Z	P53544 human herpe
40	32	61.5	759	1 SCTL_YEAST	P32784 saccharomyc
41	32	61.5	920	1 EDD_RAT	Q62671 rattus norv
42	32	61.5	993	1 VIA_TAV	P28931 tomato aspe
43	32	61.5	1377	1 RHSA_ECOLI	P16916 escherichia
44	32	61.5	1397	1 RHSC_ECOLI	P16918 escherichia
45	32	61.5	1411	1 RHSS_ECOLI	P16917 escherichia

ALIGNMENTS

RESULT 1

ID	CARB_FUSNN	STANDARD;	PRT;	1058 AA.
AC	Q8RG86;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
GN	CARB OR FN0422.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; Fusobacterium.			
OK	NCBI_TaxID=76856;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 25586;			
RX	MEDLINE=21886394; PubMed=11889109;			
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., Du Suza M., Walunas T., Pusch G., Haselkorn R., Festerstein M., Kyrpides N., Overbeek R.;			
RA	"Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";			
RL	J. Bacteriol. 184:2005-2018(2002).			
CC	- - CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.			
CC	- - COFACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	- - PATHWAY: Arginine biosynthesis; first step.			
CC	- - SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).			
CC	- - SIMILARITY: Belongs to the carb family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AE010554; AAL94625.1; ALT_INIT.			
DR	HAMAP; MF_01210; -; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; CPase_L.			
DR	InterPro; IPR005479; CPase_L_D2.			
DR	InterPro; IPR005480; CPase_L_D3.			
DR	InterPro; IPR005481; CPase_L_N.			
DR	InterPro; IPR004362; MGS_Like.			
DR	Pfam; PF00289; CPase_L_chain; 2.			
DR	Pfam; PF02786; CPase_L_D2; 2.			
DR	Pfam; PF02787; CPase_L_D3; 1.			

```

DR Pfam: PF02142; MGS; 1.
DR PRINTS; PRO0098; CFSASE.
DR TIGRFAMs; TIGR01369; CFSASE1_lrg; 1.
DR PROSITE; PS00866; CFSASE 1; 2.
DR PROSITE; PS00867; CFSASE 2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGMYS 11
Db 190 EIVPGLNYS 199

RESULT 2
ID SULH_SCHPO STANDARD; PRT; 877 AA.
AC 074377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable sulfate permease C3H7.02.
GN SPBC3H7.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Spours J., Peat N., Hayes J., Baker S., Basham D., Boman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wiltjens I., Vansireals E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler R., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53)
CC family.
CC -!- SIMILARITY: Contains 1 STAS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL031261; CAA20298.1; -.
CC PIR; T40413; T40413.
CC GenBank; SPombe; SPBC3H7.02; -.
CC InterPro; IPR002645; STAS.
CC InterPro; IPR001902; Sulph_transpt.
CC Pfam; PF01740; STAS; 1.
CC Pfam; PF00916; Sulfate_transp; 1.
CC TIGRFAMs; TIGR00815; sulp; 1.
CC PROSITE; PS01130; SLC26A; 1.
CC PROSITE; PS00801; STAS; 1.
CC Transport; Transmembrane.
KW TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
FT TRANSMEM 484 504 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 543 563 POTENTIAL.
FT DOMAIN 594 747 STAS.
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
Best Local Similarity 77.8%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGMYS 11
Db 148 VVPGMSYA 156

RESULT 3
ID Y1LK_TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN Y1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92189538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants.";
RT Virology 187:633-642(1992).
RL

```


)R InterPro; IPR02543; FtsK_SpoIIIE.

CC between the Swiss Institute of
CC the European Information Institute

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; X51435; CAA35798.1; --
 CC F1R; A34203; A34203.
 CC PDB; 3ZNF; 15-JAN-92.
 CC PDB; 4ZNF; 15-JAN-92.
 CC PDB; 1BBO; 31-OCT-93.
 CC TRANSFAC; T00497; --
 CC Genew; HGNC:4920; HIVEP1.
 CC MIM; 194540; --
 CC GO; GO:0005634; C:nucleus; TAS.
 CC GO; GO:0003677; F:DNA binding; TAS.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; zf-C2H2; 5.
 CC SMART; SM00355; Znf_C2H2; 4.
 CC PROSITE; PS0028; ZINC_FINGER_C2H2_1; 4.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Repeat; 3D-structure.
 FT ZN_FING 406 428
 FT ZN_FING 434 456
 FT ZN_FING 958 981
 FT ZN_FING 2087 2109
 FT ZN_FING 2115 2139
 FT ZN_FING 2139 2139
 FT DOMAIN 803 806
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2095 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 2405 VVPAGLTYS 2413

RESULT 6
 CY14_NEUCR
 ID_C114_NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate permease II.
 GN CYS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129256; PubMed=1825178;
 RA Ketter J.S., Jarai G., Fu Y.-H., Marzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of cys-14, the structural gene for sulfate permease II in
 RT Neurospora crassa";
 RL Biochemistry 30:1780-1787(1991).
 RN [2]
 RP PROBABLE REVISIONS
 RX MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate

RT permease II and a putative human tumour suppressor."; Trends Biochem. Sci. 19:19-19(1994).
 CC -!- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- INDUCTION: Highly expressed, but only in cells subject to sulfur limitation, and it is turned on by the positive-acting Cys-3 sulfur regulatory protein.
 CC -!- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC -!- SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53) family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 90 VVPOGMAYA 98

RESULT 7
 A10A_HUMAN
 ID_A10A_HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; Q96914;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
 DE (Aminophospholipid translocase VA).
 GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21225279; PubMed=11326269;
 RA Meguro M., Kashiwagi A., Mitsuya K., Nakao M., Kondo I., Saich S., Oshimura M.;
 RT "A novel maternally expressed gene, ATP10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome."; Nat. Genet. 28:19-20(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

```

3X MEDLINE=21313119; PubMed=11353404;
3A Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
3T "The human aminophospholipid-transferring ATPase gene ATP10C maps
3U adjacent to UBE3A and exhibits similar imprinted expression.";
3L Am. J. Hum. Genet. 68:1501-1505(2001).
3N [3]
3P SEQUENCE FROM N.A.
3C TISSUE=Skin;
3X MEDLINE=2238257; PubMed=12477932;
3A Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
3A Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
3A Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
3A Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
3A Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
3A Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
3A Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
3A Rana S.S., Locoallano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
3A Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
3A Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
3A Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
3A Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
3A Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
3A Blackley R.W., Touchman J.W., Green E.D., Dickson M.C.,
3A Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
3A Butterfield Y.S.N.F., Krzywinski M.I., Skalska U., Smallos D.E.,
3A Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
3A "Generation and initial analysis of more than 15,000 full-length
3T human and mouse cDNA sequences.";
3U Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
3N [4]
3P SEQUENCE OF 337-1499 FROM N.A.
3C TISSUE=Brain;
3X MEDLINE=36290545; PubMed=9628581;
3A Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
3A Nomura N., Ohara O.;
3T "Prediction of the coding sequences of unidentified human genes. IX.
3U The complete sequences of 100 new cDNA clones from brain which can
3L code for large proteins in vitro.";
3C DNA Res. 5:31-39(1998).
3X -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
3C -!- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
3C -!- TISSUE SPECIFICITY: Widely expressed, with highest levels in
3C kidney, followed by lung, brain, prostate, testis, ovary and
3C small intestine.
3C -!- DISEASE: Defects in ATP10A are a cause of Angelman syndrome (AS)
3C [MIM:105830]; also known as 'happy puppet syndrome'. AS is
3C characterized by features of severe motor and intellectual
3C retardation, microcephaly, ataxia, frequent jerky limb movements
3C and flapping of the arms and hands, hypotonia, hyperactivity,
3C hypopigmentation, seizures, absence of speech, frequent smiling
3C and episodes of paroxysmal laughter, and an unusual facies
3C characterized by macrostomia, a large mandible and open-mouthed
3C expression, a great propensity for protruding the tongue ('tongue
3C thrusting'), and an occipital groove.
3C -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
3C ATPases). Subfamily IV.
3C
3C This SWISS-PROT entry is copyright. It is produced through a collaboration
3C between the Swiss Institute of Bioinformatics and the EMBL outstation -
3C the European Bioinformatics Institute. There are no restrictions on its
3C use by non-profit institutions as long as its content is in no way
3C modified and this statement is not removed. Usage by and for commercial
3C entities requires a license agreement (See http://www.isb-sib.ch/announce/
3C or send an email to license@isb-sib.ch).
3C
3C EMBL; AB051358; BAB47392.1; -
3C EMBL; AY029504; AAK33100.1; -
3C EMBL; AY029487; AAK33100.1; JOINED.
3C EMBL; AY029488; AAK33100.1; JOINED.
3C EMBL; AY029489; AAK33100.1; JOINED.
3C EMBL; AY029490; AAK33100.1; JOINED.
3C EMBL; AY029491; AAK33100.1; JOINED.
3C EMBL; AY029492; AAK33100.1; JOINED.
3C
3C Query Match 65.4%; Score 34; DB 1; Length 1499;
3C Best Local Similarity 72.7%; Pred. No. 70;
3C Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
3C
3C QY 1 EEVVPXGMSVS 11
3C Db 469 EEVVPXGMSVS 479
3C
3C RESULT 8
3C RL20_MTCPU STANDARD; PRT; 116 AA.
3C ID RL20_MTCPU
3C AC Q38QV0;
3C DT 28-FEB-2003 (Rel. 41, Created)
3C DT 28-FEB-2003 (Rel. 41, Last sequence update)
3C DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE 50S ribosomal protein L20.
GN RPLT OR MYPU 2610.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP; PubMed=11353084;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis."
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -!- FUNCTION: This protein binds directly to 23S ribosomal RNA and is
CC necessary for the in vitro assembly process of the 50S ribosomal
CC subunit. It is not involved in the protein synthesizing functions
CC of that subunit (By similarity).
CC -!- SIMILARITY: Belongs to the L20P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445563; CAC13434.1; -.
DR PIR; E90544; E90544.
DR MYPULIST; MYPU_2610; -.
DR HAMAP; MF_00382; -.
DR InterPro; IPR005813; Ribosomal L20.
DR InterPro; IPR005812; Ribosomal_L20b/o.
DR Pfam; PF00453; Ribosomal_L20; I.
DR PRINTS; PR00662; RIBOSOMALL20.
DR PRODOM; PD002389; L20; I.
DR TIGRFAMs; TIGR01032; rplT bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 53.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 8.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VWPXGMSYS 11
DB 68 VRPLGMSYS 76

RESULT 9
ID YJ49 ARCFU STANDARD; PRT; 165 AA.
AC 028330.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gunn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kurlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

```

```

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Mason T.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000968; AAB89307.1; -.
DR PIR; D69493; D69493.
DR TIGR; AF1949; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 141 161
SQ SEQUENCE 165 AA; 17588 MW; B6C17054810ADB8F8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 60 BESIPDGASY 69

RESULT 10
ID Y990 CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990c.
GN C00990c.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli."
RL J. Bacteriol. 177:2396-2402(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AL139076; CAB73246.1; -
 DR EMBL; Z36940; CA85392.1; -
 DR PIR; C81374; C81374.
 DR PIR; I40758; I40758.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
 Best Local Similarity 55.6%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 ::|||
 Db 185 DFFPGMSY 193

RESULT 11

CTX3 MOUSE

ID CTX3 MOUSE STANDARD; PRT; 280 AA.
 AC Q9D387; Q9CXQ4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 SN C20ORF103
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staebli F., Suzuki R., Tonita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT Functional annotation of a full-length mouse cDNA collection.;
 RL Nature 403:685-690(2000).
 RN [2]

SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shennen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer C.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Gramwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
 RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AK014127; BAB29169.1; -
 DR EMBL; AK018222; BAB31124.1; ALT_FRAME.
 DR EMBL; BC004791; AA04791.1; -
 DR MGB; MG1:1920368; 3110035N03Rik.
 DR MGB; MG1:1923411; 6330527006Rik.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 280 PROTEIN C20ORF103 HOMOLOG.
 FT DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 236 256 POTENTIAL.
 FT DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 118 Missing (in isoform 2).
 FT FTID=VSP_003820.
 FT CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).
 FT CONFLICT 230 230 Q -> P (IN REF. 1; BAB31124).
 FT CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).
 SQ SEQUENCE 280 AA; 31721 MW; FA1D7BF9FD5CCEFCRC64;

Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSY 10
 |||||
 Db 173 VTPAGMSY 180

RESULT 12

AROA_VIBCH

ID AROA_VIBCH STANDARD; PRT; 426 AA.
 AC Q9KRB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoheximate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 enolpyruvylthiokinimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR VC1732.
 OS Vibrio cholerae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=666;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=El Tor N16961 / Serotype O1;
RX  MEDLINE=20406833; PubMed=10952301;
RA  Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA  Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA  Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA  Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA  McDonald L., Uterback T., Fleischmann R.D., Niernan W.C., White O.,
RA  Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA  Fraser C.M.;
RT  "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT  cholerae.";
RL  Nature 406:477-483(2000).
CC  -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC  phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC  -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC  sixth step.
CC  -1- SUBUNIT: Monomer (By similarity).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -1- SIMILARITY: Belongs to the EPSP synthase family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AE004251; AAF94882.1; --
CC  PIR; D82163; D82163.
CC  TIGR; VC1732; --
CC  HAMAP; MF 00210; -; 1.
CC  InterPro; IPR006264; AroA.
CC  InterPro; IPR001986; EPSP synth.
CC  Pfam; PF00275; EPSP synthase; 1.
CC  PRODOM; PD001867; EPSP synthase; 1.
CC  TIGRfam; TIGR01356; aroA; 1.
CC  PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC  PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
CC  KW Aromatic amino acid biosynthesis; Transference; Complete proteome.
CC  SEQUENCE 426 AA; 46101 MW; 38852D6483BF31C3 CRC64;
SQ  SEQUENCE 426 AA; 46101 MW; 38852D6483BF31C3 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 223 EFVIPAGQSY 232

RESULT 13
CC37 SCHFO
ID CC37 SCHDO STANDARD; PRT; 466 AA.
AC O94740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE subunit) (Cell division control protein 37).
GN CDC37 OR SPBC986.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
CX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornaby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vantreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Meestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Dominguez A., Revuelta J.L., Moreno S., Sanchez M., del Rey F., Benito J.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Forsburg S.L.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [3]
RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=22745456; PubMed=12861001;
RA Tatebe H., Shiozaki K.;
RT "Identification of cdc37 as a novel regulator of the stress-responsive
RT mitogen-activated protein kinase.";
RL Mol. Cell. Biol. 23:5132-5142(2003).
CC -1- FUNCTION: Co-chaperone that binds to numerous kinases and promotes
CC their interaction with the Hsp90 complex, resulting in
CC stabilization and promotion of their activity.
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with sty1.
CC -1- SUBCELLULAR LOCATION: Nuclear, and cytoplasmic. When in the
CC nucleus associated with chromatin.
CC -1- SIMILARITY: Belongs to the CDC37 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ132377; CAB38758.1; --
CC EMBL; AJ132376; CAB38757.1; --
CC EMBL; AL049769; CAB42371.2; --
CC PIR; T43653; T43653.
CC GenBank; SPombe; SPBC986.10; --
CC InterPro; IPR004918; Cdc37.
CC Pfam; PF03234; Cdc37; 1.
CC KW Chaperone; Cell division; Cell cycle; Nuclear protein.
CC SEQUENCE 466 AA; 52554 MW; 647236B34CABB3C5 CRC64;
SQ  SEQUENCE 466 AA; 52554 MW; 647236B34CABB3C5 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 466;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 98 DSAIPGMSY 107

RESULT 14
```

3SR2 HUMAN
 ID Q9NZM5; Q9BHC6; Q9HAX6; Q9NPP1; Q9NPR4; Q9UF12; 478 AA.
 CC GSRMZ5; Q9BHC6; Q9HAX6; Q9NPP1; Q9NPR4; Q9UF12;
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF182076; AB562873.1; -
 DR EMBL; BC004329; AAH04229.1; -
 DR EMBL; BC006311; AAH06311.1; -
 DR EMBL; BC010095; AAH10095.1; -
 DR EMBL; AF296124; AAG30413.1; -
 DR EMBL; AF296124; AAG30413.1; -
 DR EMBL; AL359335; CAB94786.1; -
 DR EMBL; AL359336; CAB94787.1; -
 DR EMBL; AL122063; CAB59242.1; -
 DR SWISS-2DRAGE; Q9NZM5; HUMAN.
 DR Genew; HGNC:4333; GLTSCR2.
 DR MIM; 605691; -
 DR GO; GO:0005622; C:intracellular; NAS.
 KW Nuclear protein; Polymorphism.
 FT VARIANT 389 389
 FT R -> Q.
 FT /FTID=VAR_011486.
 FT CONFLICT 4 6
 FT GGS -> HEG (IN REF. 2; AAH04229).
 FT CONFLICT 9 9
 FT G -> R (IN REF. 3).
 FT CONFLICT 146 191
 FT RRKEQWELKAGQGLFRRVRAQARLNPSATRAKPGQD
 FT TVERP -> SGRSSYGRSWPSSPSSGAGSPVAQPCN
 FT KGNPAPGHR1AA (IN REF. 3).
 FT CONFLICT 198 215
 FT SDNPLDRPLVGQDEFFLE -> LNNPDKPVVWPGCLFPG
 FT (IN REF. 3).
 FT CONFLICT 235 235
 FT A -> S (IN REF. 2; AAH04229).
 FT CONFLICT 417 417
 FT D -> H (IN REF. 3).
 FT CONFLICT 433 477
 FT PEGNILDRLFKFQRRNMIEPRERAKFRKVKVLEKRAF
 FT REIQ -> VLVTVSCRGAPCVMTPLLPVPRGYGRHHGCP
 FT WAGPVGMPERG (IN REF. 5).
 FT CONFLICT 434 478
 FT EGNILDRFKSFQRRNMIEPRERAKFRKVKVLEKRAF
 FT ETQL -> RQSHSFETGSAFPGGI (IN REF. 3).
 SQ SEQUENCE 478 AA; 54417 MW; 7F18923E348CB52B CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 478;
 Best Local Similarity 60.0%; Pred. NO. 35;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPXGMSYS 11
 DB 239 EVAPAGASTN 248
 RESULT 15
 BCN5 CLOPE
 ID BCN5 CLOPE STANDARD; PRT; 890 AA.
 AC P08636;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DB Bacteriocin BCN5.
 GN BCN.
 OS Clostridium perfringens.
 OG Plasmid pIP404.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50;
 RA MEDLINE=88336297; PubMed=2901768;
 RA Garner T., Cole S.T.;
 RT "Complete nucleotide sequence and genetic organization of the
 RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
 RL Plasmid 19:134-150(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50;
 RA MEDLINE=87057020; PubMed=2877971;
 RA Garner T., Cole S.T.;
 RT "Characterization of a bacteriocinogenic plasmid from Clostridium

```
RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene."
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro."
RL Mol. Microbiol. 2:607-614(1988).
CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -1- INDUCTION: By UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -.
DR EMBL; M32882; AAA98249.1; -.
DR PIR; A30481; A30481.
DR InterPro; IPR000834; Peptidase_M14.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR SMART; SM00287; SH3b; 3.
DR Antibiotic; Bacteriocin; Plasmid.
KW DOMAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;
SQ -----
Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 EVVPXGMSY 10
Db 170 EVVPGGFTY 178
|||||:|
| | | | |
```

Search completed: June 3, 2004, 11:49:50
Job time : 6.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-5

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP arChaea.*
- 2: SP bacteria.*
- 3: SP fungi.*
- 4: SP human.*
- 5: SP invertebrate.*
- 6: SP mammal.*
- 7: SP mhc.*
- 8: SP organelle.*
- 9: SP phage.*
- 10: SP plant.*
- 11: SP rodent.*
- 12: SP virus.*
- 13: SP vertebrate.*
- 14: SP unclassified.*
- 15: SP virus.*
- 16: SP bacteriap.*
- 17: SP archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	75.0	1044	Q8DIH0	Q8diH0 synechococ
2	38	73.1	344	Q815A7	Q815a7 bacillus ce
3	38	73.1	387	Q98FX1	Q98fx1 rhizobium 1
4	38	73.1	3472	1 Q74056	Q74056 cenarchaeum
5	37	71.2	840	Q9URY8	Q9ury8 schizosacch
6	37	71.2	1123	Q8EWD4	Q8ewd4 mycoplasma
7	36	69.2	471	Q8R126	Q8r126 mus musculu
8	36	69.2	484	Q8VD18	Q8vd18 mus musculu
9	36	69.2	484	Q8BTX4	Q8btx4 mus musculu
10	36	69.2	484	Q8BK35	Q8bk35 mus musculu
11	36	69.2	559	Q839T9	Q839t9 enterococcu
12	36	69.2	1399	Q889X7	Q889x7 pseudomonas
13	35	67.3	225	Q40129	Q40129 lycopersico
14	35	67.3	245	Q7V6Q4	Q7v6q4 prochloroco
15	35	67.3	425	Q9XVK4	Q9xvk4 caenorhabd
16	35	67.3	495	Q8CLD7	Q8clD7 mus musculu

17	35	67.3	555	4	Q7Z6R0	Q7z6r0 homo sapien
18	35	67.3	583	5	Q9BH83	Q9bh83 plasmodium
19	35	67.3	583	5	Q9BHA5	Q9bha5 plasmodium
20	35	67.3	583	5	Q815S7	Q815s7 plasmodium
21	35	67.3	670	11	Q01487	Q01487 rattus ratt
22	35	67.3	747	16	Q8PMI6	Q8pmi6 xanthomonas
23	35	67.3	747	16	Q8PAT2	Q8pat2 xanthomonas
24	35	67.3	749	16	Q8PDM6	Q8pdm6 xylella fas
25	35	67.3	1902	4	Q14122	Q14122 homo sapien
26	34	65.4	156	3	Q12479	Q12479 saccharomyc
27	34	65.4	219	17	Q971S2	Q971s2 sulfolobus
28	34	65.4	252	17	Q28342	Q28342 archaeoglob
29	34	65.4	290	4	Q96NU1	Q96nu1 homo sapien
30	34	65.4	387	16	Q92MD6	Q92md6 rhizobium m
31	34	65.4	489	4	Q8IYM3	Q8iym3 homo sapien
32	34	65.4	541	16	Q98BP5	Q98bp5 rhizobium 1
33	34	65.4	544	16	Q9PQD2	Q9pqd2 ureaplasma
34	34	65.4	731	16	Q7UWU7	Q7uwu7 rhodopirell
35	34	65.4	842	3	Q9URR4	Q9urra4 penicillium
36	34	65.4	899	16	Q8G4I5	Q8g4i5 bifidobacte
37	34	65.4	1049	16	Q8XT05	Q8xt05 raistonia s
38	34	65.4	1400	16	Q87KQ5	Q87kq5 vibrio para
39	33	63.5	143	17	Q8TX62	Q8tx62 methanopyru
40	33	63.5	166	16	Q8PPP5	Q8ppp5 xanthomonas
41	33	63.5	193	2	Q8VUA8	Q8vua8 lactococcus
42	33	63.5	208	2	Q8XTQ4	Q8xtq4 candidatus
43	33	63.5	209	16	Q8RE56	Q8re56 fusobacteri
44	33	63.5	251	13	Q7SY67	Q7sy67 xenopus lae
45	33	63.5	282	16	Q7U552	Q7u552 synechococc

ALIGNMENTS

RESULT 1

Q8DIH0	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
AC	Q8DIH0	01-MAR-2003 (T-EMBLrel. 23, Created)	
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)	
DT	01-JUN-2003	(T-EMBLrel. 24, Last annotation update)	
DE	Multidrug efflux transporter.		
GN	TLL1618.		
OS	Synechococcus elongatus (Thermosynechococcus elongatus).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI_TaxID=32046;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-1;		
RX	MEDLINE=2225144; PubMed=12240834;		
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,		
RA	Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,		
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of the thermophilic cyanobacterium		
RL	Thermosynechococcus elongatus BP-1.";		
RL	DNA Res. 9:123-130(2002).		
DR	EMBL; AP005374; BAC09170.1; ..		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0005215; P:transporter activity; IEA.		
DR	GO; GO:0006810; P:transport; IEA.		
DR	InterPro; IPR01036; Acrlivin_res.		
DR	Pfam; PF00873; ACR_tran; 1.		
DR	PRINTS; PR00702; ACRIFLAVINRP.		
DR	TIGRFAMs; TIGR00915; 2A0602; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;		

Query Match 75.0%; Score 39; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 EVVVPXGMSYS 11
Db 843 EVLVNGIGYS 853

RESULT 2
Q815A7 PRELIMINARY; PRT; 344 AA.
AC Q815A7;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BC5259.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyripides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003)
DR EMBL; AE017015; AAP12123.1; -.
DR InterPro; IPR000437; Prok_lipoprot.S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 344 AA; 38539 MW; C55268ACB7225995 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 344;
Best Local Similarity 60.0%; Pred. NO. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
Db 152 EETAPLGSLY 161

RESULT 3
Q98FX1 PRELIMINARY; PRT; 387 AA.
AC Q98FX1;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hippurate hydrolase.
GN MLR3583.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50445.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase_M20.

DR Pfam; PF01546; Peptidase_M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
Best Local Similarity 60.0%; Pred. NO. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
Db 367 DEAIPEGMSY 376

RESULT 4
O74056 PRELIMINARY; PRT; 3472 AA.
AC O74056;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
OC Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B;
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL; AF083072; AAC62699.1; -.
DR FIR; T31308; T31308.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD_transp.
DR InterPro; IPR001680; WD40.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Hypothetical protein.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
Db 2294 EDVIPRGISFS 2304

RESULT 5
Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972h-;
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;

```

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

EMBL; AL32779; CAB60015.1; -;
PIR; T39116; T39116.
GeneDB SPombe; SPAC969.05G; -;
GO; GO:0016020; C:membrane; IEA.
GO; GO:0008271; F:sulfate porter activity; IEA.
GO; GO:0008272; P:sulfate transport; IEA.
InterPro; IPR002645; SPAS.
InterPro; IPR001902; Sulph_transpt.
Pfam; PF01740; STAS; 1.
Pfam; PF00916; Sulfate transp; 1.
TIGRFAMs; TIGR00815; sulp; 1.
PROSITE; PS50801; STAS; 1.
SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

3 VVPXGMSYS 11
||| |||||
135 VVPQMSYA 143

RESULT 6

ID Q8EWD4 PRELIMINARY; PRT; 1123 AA.
AC Q8EWD4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
YE MYPE 2560 paralogs, 57%.
IN MYPE2710.
NS Mycoplasma penetrans.
NC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
SA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
SA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
RL EMBL; AP004171; BAC44062.1; -;
JR InterPro; IPR008985; ConA like lec gl.
JR InterPro; IPR007326; Lipoprotein_17.
JR Pfam; PF04200; Lipoprotein_17; 3.
KW Complete proteome.
SQ SEQUENCE 1123 AA; 123636 MW; A4D70730E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
Best Local Similarity 70.0%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

2 EVVPXGMSYS 11
||| |||||
658 EVVPMGLSYS 667

RESULT 7

ID Q8R126 PRELIMINARY; PRT; 471 AA.
AC Q8R126;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
ZN GLTSCR2.
XS Mus musculus (Mouse).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OK NCBI_TaxID=10090;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025810; AAH25810.1; -;
DR MGD; MGI:2154441; GLTscr2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

2 EVVPXGMSYS 11
||| |||||
226 EVIPAGASYN 235

RESULT 8

ID Q8VD18 PRELIMINARY; PRT; 484 AA.
AC Q8VD18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AW536441.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary Gland;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017637; AAH17637.1; -;
DR MGD; MGI:2154441; GLTscr2.
SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

2 EVVPXGMSYS 11
||| |||||
239 EVIPAGASYN 248

RESULT 9

ID Q8BTX4 PRELIMINARY; PRT; 484 AA.
AC Q8BTX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```

RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK088461; BAC0367.1; -.
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EECAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 10
Q8BK35 PRELIMINARY; PRT; 484 AA.
AC Q8BK35; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
GN Gltscr2.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK077341; BAC36760.1; -.
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55792 MW; EB67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 11
Q839T9 PRELIMINARY; PRT; 559 AA.
AC Q839T9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phorbolone binding protein, putative.
GN EF0063.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
CX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Sehadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;

"Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis."
RL Science 299:2071-2074(2003).
DR EMBL; AB016947; AAO79943.1; -.
DR TIGR; EF0063; -.
DR GO; GO:0005215; P-transporter activity; IEA.
DR GO; GO:000810; P-transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 559 AA; 61476 MW; CC15418D33D53DE7 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 559;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 351 LIPEGMSYS 359

RESULT 12
Q889X7 PRELIMINARY; PRT; 1399 AA.
AC Q889X7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA-directed RNA polymerase, beta' subunit.
GN RPOC OR PSPT00620.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
CX Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Barry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengat J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RA "Complete sequence of Pseudomonas syringae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016858; AAO54162.1; -.
DR TIGR; PSPT00620; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008350; P:transcription; IEA.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR DNA-directed RNA polymerase; Complete proteome.
SQ SEQUENCE 1399 AA; 154733 MW; 26178BD653102ADE CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1399;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 10
Db 581 QVVPAGLSY 589

```

```

RESULT 13
D Q40129 PRELIMINARY; PRT; 225 AA.
C Q40129
T 01-NOV-1996 (TrEMBLrel. 01, Created)
T 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Hypothetical protein precursor.
S Lycopersicon esculentum (Tomato).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C Lamnids; Solanales; Solanaceae; Solanum.
X NCBI_TaxID=4081;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=VF36; TISSUE=Plant;
C MEDLINE=95375233; PubMed=7647301;
A Mulligan S.B., Gasser C.S.;
T "Nature and regulation of pistil-expressed genes in tomato.";
L Plant Mol. Biol. 28:691-711(1995).
R ENBL; U20592; AAA80497.1; -.
R PIR; S57810; S57810.
R GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
R InterPro; IPR002160; Kunitz_legume.
R Pfam; PF00197; Kunitz_legume; 1.
R PRINTS; PR00291; KUNITZINHBTR.
R ProDom; PD000891; Kunitz_legume; 1.
R SMART; SM00452; ST1; 1.
R PROSITE; PS00283; SOYBEAN KUNITZ; 1.
M Hypothetical protein; Signal. POTENTIAL.
T SIGNAL 1 20
T CHAIN 21 225
T Q SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Y 1 EVVVPXGMSYS 11
b :|||||:|:|:
32 DEVVPNGKTYA 42

RESULT 14
D Q7V6Q4 PRELIMINARY; PRT; 245 AA.
C Q7V6Q4
T 01-OCT-2003 (TrEMBLrel. 25, Created)
T 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Phospholipid and glycerol acyltransferase (From 'motifs_6.msaf').
N PWT1092.
S Prochlorococcus marinus (strain MIT 9313).
C Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
C Prochlorococcus.
X NCBI_TaxID=74547;
N [1]
P SEQUENCE FROM N.A.
C MEDLINE=22825698; PubMed=12917642;
A Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
A Algren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
A Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
A Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
A Webb E.A., Zinser E.R., Chisholm S.W.;
T "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
T niche differentiation.";
L Nature 424:1042-1047(2003).
R ENBL; BX572098; CAE21267.1; -.
W Acyltransferase; Transferase; Complete proteome.
Q SEQUENCE 245 AA; 26907 MW; 106F7C4CB2C6427 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 245;

```

```

Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGMSYS 11
Db :|||||:|:|:
179 QVVPVGLGYS 188

RESULT 15
D Q9XVK4 PRELIMINARY; PRT; 425 AA.
C Q9XVK4
T 01-NOV-1999 (TrEMBLrel. 12, Created)
T 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
N [1]
P SEQUENCE FROM N.A.
C Percy C.M.;
R Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RL [2]
R SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR ENBL; Z81109; CAB03241.1; -.
DR PIR; T24111; T24111.
DR WormPep; R10D12.10; CE12690.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004672; P:protein kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_Kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

```

```

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 BEVVPXGMSY 10
Db :|||||:|:|:
335 EQIVFGGLQY 344

```

```

Search completed: June 3, 2004, 11:57:30
Job time : 30.8667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	5	ABB80525 Hepatitis
2	50	96.2	11	5	ABB80521 Hepatitis
3	50	96.2	11	5	ABB80522 Hepatitis
4	50	96.2	11	5	ABB80566 Hepatitis
5	50	96.2	11	5	ABB80563 Hepatitis
6	50	96.2	11	5	ABB80565 Hepatitis
7	50	96.2	11	5	ABB80567 Hepatitis
8	50	96.2	11	5	ABB80559 Hepatitis
9	50	96.2	11	5	ABB80526 Hepatitis
10	50	96.2	11	5	ABB80564 Hepatitis
11	50	96.2	11	5	ABB80568 Hepatitis
12	46	88.5	11	5	ABB80561 Hepatitis
13	46	88.5	11	5	ABB80524 Hepatitis
14	46	88.5	11	5	ABB80529 Hepatitis
15	46	88.5	11	5	ABB80528 Hepatitis
16	46	88.5	11	5	ABB80562 Hepatitis
17	45	86.5	11	5	ABB80523 Hepatitis
18	45	86.5	11	5	ABB80536 Hepatitis
19	45	86.5	11	5	ABB80558 Hepatitis
20	45	86.5	11	5	ABB80560 Hepatitis
21	45	86.5	11	5	ABB80527 Hepatitis
22	45	86.5	11	5	ABB80535 Hepatitis
23	45	86.5	11	5	ABB80540 Hepatitis
24	45	86.5	11	5	ABB80539 Hepatitis
25	44	84.6	11	5	ABB80549 Hepatitis

26	44	84.6	11	5	ABB80544	Hepatitis
27	44	84.6	11	5	ABB80553	Hepatitis
28	44	84.6	11	5	ABB80552	Hepatitis
29	44	84.6	11	5	ABB80545	Hepatitis
30	42	80.8	11	5	ABB80530	Hepatitis
31	41	78.8	11	5	ABB80542	Hepatitis
32	41	78.8	11	5	ABB80543	Hepatitis
33	41	78.8	11	5	ABB80538	Hepatitis
34	40	76.9	11	5	ABB80548	Hepatitis
35	40	76.9	11	5	ABB80547	Hepatitis
36	40	76.9	11	5	ABB80556	Hepatitis
37	40	76.9	11	5	ABB80557	Hepatitis
38	40	76.9	11	5	ABB80537	Hepatitis
39	40	76.9	11	5	ABB80551	Hepatitis
40	40	76.9	11	5	ABB80541	Hepatitis
41	40	76.9	20	2	AAU76810	Hepatitis
42	40	76.9	1022	4	ABG03621	Novel hum
43	40	76.9	1022	4	ABG08173	Novel hum
44	40	76.9	1022	4	ABG05826	Novel hum
45	39	75.0	11	5	ABB80546	Hepatitis

ALIGNMENTS

RESULT 1

ABB80525	ID	ABB80525	standard; peptide; 11 AA.
AC	ABB80525;		
XX	08-OCT-2002	(first entry)	
DT	Hepatitis C virus NS3/NS4a	serine protease inhibitor peptide #5.	
DE	Hepatitis C virus; HCV;	serine protease; inhibitor; alpha-ketoamide;	
XX	virucide.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
FT	Modified-site 1	/note= "N-terminal acetyl"	
FT	Modified-site 6	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"	
FT	Misc-difference 8	/note= "D-form residue"	
FT	Modified-site 11	/note= "C-terminal amide"	
XX	WO200208251-A2.		
XX	31-JAN-2002.		
XX	19-JUL-2001;	2001WO-US023169.	
XX	21-JUL-2000;	2000US-0220101P.	
XX	(CORV-) CORVAS	INT INC.	
XX	Lim-Wilby M, Levy OE, Brunck TK;		
XX	WPI; 2002-361643/39.		
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.		
XX	Claim 17; Page 64; 69pp; English.		
XX	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the		

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
 |||||
 DB 1 EEVVPXGMSYS 11

RESULT 2

ABB80521
 ID ABB80521 standard; peptide; 11 AA.

XX AC ABB80521;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

PN 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEVVPXGMSYS 11
 |||||
 DB 1 EEVVPXGMSYS 11

RESULT 3

ABB80522
 ID ABB80522 standard; peptide; 11 AA.

XX AC ABB80522;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

PN 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
 |||||
 DB 1 EEVVPXGMSYS 11

RESULT 4

```

ABB80566
ID ABB80566 standard; peptide; 11 AA.
CX
CX ABB80566;
CX
CX 08-OCT-2002 (first entry)
CX
CX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
CX
CX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
CX virucide.
CX
CX Synthetic.
CX
CX Key Location/Qualifiers
CX Modified-site 1 /note= "N-terminal acetyl"
CX
CX Modified-site 6 /note= "2-aminoisobutyryl carbonyl residue forming a keto
CX -amide linkage with residue 7"
CX
CX Modified-site 11 /note= "C-terminal amide"
CX
CX WO200208251-A2.
CX
CX 31-JAN-2002.
CX
CX 19-JUL-2001; 2001WO-US023169.
CX
CX 21-JUL-2000; 2000US-0220101P.
CX
CX (CORV-) CORVAS INT INC.
CX
CX Lim-Wilby M, Levy OE, Brunck TK;
CX WPI; 2002-361643/39.
CX
CX Novel peptide compound having hepatitis C virus protease inhibitory
CX activity useful for treating disorders associated with hepatitis C virus
CX protease.
CX
CX Claim 17; Page 65; 69pp; English.
CX
CX The sequence represents a peptide compound of the invention having
CX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CX invention are alpha-ketoamide peptide analogues. The peptides have
CX virucide activity, and are useful for treating and in the manufacture of
CX a medicament to treat disorders associated with HCV protease. A
CX pharmaceutical composition comprising the peptide as an active ingredient
CX is useful for treating disorders associated with hepatitis C virus
CX
CX Sequence 11 AA;
CX
CX Query Match 96.2%; Score 50; DB 5; Length 11;
CX Best Local Similarity 100.0%; Pred. No. 0.002;
CX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CX
CX Qy 1 EEVVPXGMSYS 11
CX |||||
CX Db 1 EEVVPXGMSYS 11
CX
CX RESULT 6
CX ABB80565
CX ID ABB80565 standard; peptide; 11 AA.
CX
CX AC ABB80565;
CX
CX 08-OCT-2002 (first entry)
CX
CX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
CX
CX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
CX virucide.
CX
CX Synthetic.
CX
CX Key Location/Qualifiers
CX Modified-site 1 /note= "N-terminal acetyl"
CX
CX Modified-site 6 /note= "Norleucyl carbonyl forming keto-amide linkage
CX with residue 7"
CX
CX

```


FT Modified-site 11 /note= "C-terminal amide"
FT XX WO200208251-A2.
PN XX
XX
XX
PD 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX PF
XX PR 21-JUL-2000; 2000US-0220101P.
XX
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
PT
XX Claim 17; Page 65; 69pp; English.
PS
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 7
ABB80567
ID ABB80567 standard; peptide; 11 AA.
XX AC ABB80567;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT Modified-site 8 residue 7"
FT FT Modified-site 8 /note= "Oxymethionine"
FT FT Misc-difference 8 /note= "D-form residue"
FT FT Modified-site 11 /note= "C-terminal amide"
FT XX
XX WO200208251-A2.
XX PD 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX XX
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX XX WPI; 2002-361643/39.
XX DR

XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 8
ABB80559
ID ABB80559 standard; peptide; 11 AA.
XX AC ABB80559;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT Modified-site 8 residue 7"
FT FT Modified-site 8 /note= "Oxymethionine"
FT FT Misc-difference 8 /note= "D-form residue"
FT FT Modified-site 11 /note= "C-terminal amide"
FT XX
XX WO200208251-A2.
XX PD 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX XX
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX XX WPI; 2002-361643/39.
XX DR

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 65; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 YY 1 BEVVPXGMSYS 11
 |||||
 YY 1 BEVVPXGMSYS 11
 |||||
 RESULT 9
 ABB80526
 ID ABB80526 standard; peptide; 11 AA.
 CC ABB80526;
 XX 08-OCT-2002 (first entry)
 CC Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
 CC Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CC virucide.
 CC Synthetic.
 CC Key Location/Qualifiers
 CC Modified-site 1 /note= "N-terminal acetyl"
 CC Modified-site 6 /note= "N-terminal acetyl"
 CC Modified-site 6 /note= "N-terminal acetyl"
 CC Modified-site 11 /note= "Leucyl carbonyl forming keto-amide linkage with
 CC Modified-site 11 residue 7"
 CC Modified-site 11 /note= "C-terminal amide"
 CC WO200208251-A2.
 CC 31-JAN-2002.
 CC 19-JUL-2001; 2001WO-US023169.
 CC 21-JUL-2000; 2000US-0220101P.
 CC (CORV-) CORVAS INT INC.
 CC Lim-Wilby M, Levy OE, Brunck TK;
 CC WPI; 2002-361643/39.
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 CC activity useful for treating disorders associated with hepatitis C virus
 CC protease.
 CC Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 |||||
 DB 1 BEVVPXGMSYS 11
 |||||
 RESULT 10
 ABB80564
 ID ABB80564 standard; peptide; 11 AA.
 CC ABB80564;
 XX 08-OCT-2002 (first entry)
 CC Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
 CC Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CC virucide.
 CC Synthetic.
 CC Key Location/Qualifiers
 CC Modified-site 1 /note= "N-terminal acetyl"
 CC Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with
 CC Modified-site 11 residue 7"
 CC Modified-site 11 /note= "C-terminal amide"
 CC WO200208251-A2.
 CC 31-JAN-2002.
 CC 19-JUL-2001; 2001WO-US023169.
 CC 21-JUL-2000; 2000US-0220101P.
 CC (CORV-) CORVAS INT INC.
 CC Lim-Wilby M, Levy OE, Brunck TK;
 CC WPI; 2002-361643/39.
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 CC activity useful for treating disorders associated with hepatitis C virus
 CC protease.
 CC Claim 17; Page 65; 69pp; English.
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 11

ABB80568
 ID ABB80568 standard; peptide; 11 AA.
 XX AC ABB80568;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
 FT a keto-amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 12

ABB80561
 ID ABB80561 standard; peptide; 11 AA.
 XX AC ABB80561;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMDYS 11

RESULT 13

ABB80524
 ID ABB80524 standard; peptide; 11 AA.
 XX AC ABB80524;
 XX

XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 5 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 9 /note= "D-form residue"
XX Modified-site 11 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
XX Query Match 88.5%; Score 46; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.013;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EVVVPXGMSYS 11
XX 1 EVVVPXGMDYS 11
XX RESULT 14
XX ABB80529
XX ID ABB80529 standard; peptide; 11 AA.
XX AC ABB80529;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.

XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 8 /note= "D-form residue"
XX Misc-difference 9 /note= "D-form residue"
XX Modified-site 11 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
XX Query Match 88.5%; Score 46; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.013;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EVVVPXGMSYS 11
XX 1 EVVVPXGMDYS 11
XX RESULT 15
XX ABB80528
XX ID ABB80528 standard; peptide; 11 AA.
XX AC ABB80528;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"

```

FT      residue 7"
FT      Misc-difference 8
FT      /note= "D-form residue"
FT      Modified-site 11
FT      /note= "C-terminal amide"
XX
FN      WO200208251-A2.
XX
XX      31-JAN-2002.
XX
XX      19-JUL-2001; 2001WO-US023169.
XX
XX      21-JUL-2000; 2000US-0220101P.
XX
XX      (CORV-) CORVAS INT INC.
XX
XX      Lim-Wilby M, Levy OE, Brunck TK;
XX
XX      WPI; 2002-361643/39.
XX
XX      Novel peptide compound having hepatitis C virus protease inhibitory
PT      activity useful for treating disorders associated with hepatitis C virus
PT      protease.
XX
XX      Claim 17; Page 64; 69pp; English.
XX
XX      The sequence represents a peptide compound of the invention having
CC      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC      invention are alpha-ketoamide peptide analogues. The peptides have
CC      virucide activity, and are useful for treating and in the manufacture of
CC      a medicament to treat disorders associated with HCV protease. A
CC      pharmaceutical composition comprising the peptide as an active ingredient
CC      is useful for treating disorders associated with hepatitis C virus
XX
XX      Sequence 11 AA;
XX
XX      Query Match      88.5%; Score 46; DB 5; Length 11;
XX      Best Local Similarity 90.9%; Pred. No. 0.013;
XX      Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX      Qy      1 EEVVPXGMSYS 11
XX      Db      1 EEVVPXGMDYS 11
XX

```

Search completed: June 3, 2004, 11:48:22
 Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-6
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	4	US-09-408-020-4
2	36	69.2	382	4	US-09-134-000C-3738
3	34	65.4	947	4	US-09-228-986-73
4	34	65.4	1191	4	US-09-540-236-2902
5	34	65.4	1407	4	US-09-328-352-7885
6	33	63.5	12	4	US-09-760-946-2
7	33	63.5	12	4	US-09-760-946-3
8	33	63.5	45	2	US-08-637-7598-236
9	33	63.5	45	3	US-08-871-355A-236
10	33	63.5	45	4	US-09-201-945-236
11	33	63.5	65	6	5177197-51
12	33	63.5	410	6	5177197-1
13	33	63.5	1394	6	5177197-30
14	32	61.5	10	3	US-09-357-952-66
15	32	61.5	10	4	US-09-521-650-66
16	32	61.5	10	4	US-09-168-888-66
17	32	61.5	102	2	US-08-580-988A-23
18	32	61.5	152	2	US-08-460-694-4
19	32	61.5	152	3	US-08-460-744-4
20	32	61.5	152	3	US-07-667-711B-4
21	32	61.5	173	1	US-08-193-977-7
22	32	61.5	189	2	US-08-454-517-21
23	32	61.5	189	2	US-08-246-361A-21
24	32	61.5	189	3	US-08-463-772-21
25	32	61.5	189	5	PCT-US93-05000-21
26	32	61.5	236	2	US-08-464-517-22
27	32	61.5	236	2	US-08-246-361A-22

Sequence 22, Appl
Sequence 22, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 8, Appl
Sequence 2, Appl
Sequence 19, Appl
Sequence 20, Appl
Sequence 19, Appl

Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOF.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 73.1%; Score 38; DB 4; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 2294 EDVIPRGISFS 2304

RESULT 2
US-09-134-000C-3738
; Sequence 3738, Application US/09134000C
; Patent No. 8617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032756-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3738
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Enterococcus faecalis

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (327)..(328)
; OTHER INFORMATION: Amino acids 327 & 328 are Xaa wherein Xaa = any amino acid.
US-09-134-000C-3738

Query Match      69.2%; Score 36; DB 4; Length 382;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 332 LIEPGMSYS 340

RESULT 3
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match      65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 686 VMPGMSYS 694

RESULT 4
US-09-540-236-2902
; Sequence 2902, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2902
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: M.catarrihalis
US-09-540-236-2902

Query Match      65.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 10
Db 783 EILPVGMAY 791

RESULT 5
US-09-328-352-7885

```

```

; Sequence 7885, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match      65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 10
Db 596 EVVPEGLSP 604

RESULT 6
US-09-760-946-2
; Sequence 2, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghio, Elise
; APPLICANT: Goudreau, Nathalie
; APPLICANT: Halmos, Teddy
; APPLICANT: Llinas-Brunet, Montse
; TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
; FILE REFERENCE: 13/076-1-C1
; CURRENT APPLICATION NUMBER: US/09/760,946
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/542,675
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/128,011
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Substrate for recombinant HCV NS3 protease radiometric assay
US-09-760-946-2

Query Match      63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVPXGMSYS 11
Db 1 DDIVFCMSYT 11

RESULT 7
US-09-760-946-3
; Sequence 3, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghio, Elise

```

APPLICANT: Goudreau, Nathalie
APPLICANT: Halmos, Teddy
APPLICANT: Llinas-Brinet, Montse
TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
FILE REFERENCE: 13/076-1-Cl
CURRENT APPLICATION NUMBER: US/09/760,946
CURRENT FILING DATE: 2001-08-23
PRIOR APPLICATION NUMBER: US 09/542,675
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: US 60/128,011
PRIOR FILING DATE: 1999-04-06
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tracer for NS3 protease assay
NAME/KEY: MOD_RES
LOCATION: (1)
OTHER INFORMATION: Asp at position 1 is biotinylated
NAME/KEY: MOD_RES
LOCATION: (10)
OTHER INFORMATION: Tyr at position 10 is iodinated with I-125
US-09-760-946-3

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 DDIVPCSMYSY 11

RESULT 8
US-08-637-759B-236
Sequence 236, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
HYPOTHETICAL: NO

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 DDIVPCSMYSY 11

SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1 BEISPLGWSY 10

RESULT 9
US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1 BEISPLGWSY 10

RESULT 10
 US-09-201-945-236
 ; Sequence 236, Application US/09201945
 ; Patent No. 6342215
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/201,945
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/637,759
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: RPLMS 101
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (404) 873-8794
 ; TELEFAX: (404) 873-8795
 ; INFORMATION FOR SEQ ID NO: 236:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 45 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: NO
 ; US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
 Best Local Similarity 60.0%; Pred. No. 7.8;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
 Db 1 BEISPLGWSY 10

RESULT 11
 5177197-51
 ; Patent No. 5177197
 ; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
 ; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESON-WELSH,
 ; LENA; HELDIN, CARL-HENRIK
 ; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
 ; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
 ; NUMBER OF SEQUENCES: 53
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/487,343
 ; FILING DATE: 27-FEB-1990
 ; SEQ ID NO: 51:
 ; LENGTH: 65
 ; 5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
 Best Local Similarity 45.5%; Pred. No. 12;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 Db 52 KEICPGMGYT 62
 RESULT 12
 5177197-1
 ; Patent No. 5177197
 ; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
 ; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESON-WELSH,
 ; LENA; HELDIN, CARL-HENRIK
 ; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
 ; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
 ; NUMBER OF SEQUENCES: 53
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/487,343
 ; FILING DATE: 27-FEB-1990
 ; SEQ ID NO: 1:
 ; LENGTH: 410
 ; 5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
 Best Local Similarity 45.5%; Pred. No. 97;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
 Db 399 KEICPGMGYT 409

RESULT 13
 5177197-30
 ; Patent No. 5177197
 ; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
 ; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESON-WELSH,
 ; LENA; HELDIN, CARL-HENRIK
 ; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
 ; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
 ; NUMBER OF SEQUENCES: 53
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/487,343
 ; FILING DATE: 27-FEB-1990
 ; SEQ ID NO: 30:
 ; LENGTH: 1394
 ; 5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
 Best Local Similarity 45.5%; Pred. No. 3.9e+02;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
 Db 399 KEICPGMGYT 409

RESULT 14
 US-09-357-952-66
 ; Sequence 66, Application US/09357952
 ; Patent No. 6248904
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhang, Han-Zhong
 ; APPLICANT: Cai, Sui Xiong
 ; APPLICANT: Drewe, John A.
 ; APPLICANT: Yang, Wu
 ; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whole
 ; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Protease
 ; FILE REFERENCE: 1735.0030001
 ; CURRENT APPLICATION NUMBER: US/09/357,952
 ; CURRENT FILING DATE: 1999-07-21
 ; EARLIER APPLICATION NUMBER: US 60/093,642

; EARLIER FILING DATE: 21-JUL-1998
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
JS-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DY 1 BEVVPXGMSY 10
:::| | |
DB 1 DDIVPCMSY 10

RESULT 15
US-09-521-650-66
; Sequence 66, Application US/09521650
; Patent No. 6335429
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.
; APPLICANT: Zhang, Han-Zhong
; TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1735.0290002
; CURRENT APPLICATION NUMBER: US/09/521.650
; CURRENT FILING DATE: 2000-03-09
; EARLIER APPLICATION NUMBER: 09/169,888
; EARLIER FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 60/061,582
; EARLIER FILING DATE: 1997-10-10
; EARLIER APPLICATION NUMBER: US 09/033,661
; EARLIER FILING DATE: 1998-03-03
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DY 1 BEVVPXGMSY 10
:::| | |
DB 1 DDIVPCMSY 10

Search completed: June 3, 2004, 12:03:06
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-6
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1155919 seqs, 281338677 residues

total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/1/pubaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/1/pubaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/1/pubaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	12	US-09-909-164-5
2	50	96.2	11	12	US-09-909-164-6
3	50	96.2	11	12	US-09-909-164-9
4	50	96.2	11	12	US-09-909-164-10
5	50	96.2	11	12	US-09-909-164-47
6	50	96.2	11	12	US-09-909-164-48
7	50	96.2	11	12	US-09-909-164-49
8	50	96.2	11	12	US-09-909-164-50
9	50	96.2	11	12	US-09-909-164-51
10	50	96.2	11	12	US-09-909-164-52
11	46	88.5	11	12	US-09-909-164-8
12	46	88.5	11	12	US-09-909-164-12
13	46	88.5	11	12	US-09-909-164-13
14	45	86.5	11	12	US-09-909-164-7
15	45	86.5	11	12	US-09-909-164-11

16	45	86.5	11	12	US-09-909-164-19	Sequence 19, Appl
17	45	86.5	11	12	US-09-909-164-20	Sequence 20, Appl
18	45	86.5	11	12	US-09-909-164-23	Sequence 23, Appl
19	45	86.5	11	12	US-09-909-164-24	Sequence 24, Appl
20	44	84.6	11	12	US-09-909-164-28	Sequence 28, Appl
21	44	84.6	11	12	US-09-909-164-29	Sequence 29, Appl
22	44	84.6	11	12	US-09-909-164-33	Sequence 33, Appl
23	44	84.6	11	12	US-09-909-164-36	Sequence 36, Appl
24	44	84.6	11	12	US-09-909-164-37	Sequence 37, Appl
25	44	84.6	11	12	US-09-909-164-43	Sequence 43, Appl
26	44	84.6	11	12	US-09-909-164-14	Sequence 14, Appl
27	41	78.8	11	12	US-09-909-164-22	Sequence 22, Appl
28	41	78.8	11	12	US-09-909-164-26	Sequence 26, Appl
29	41	78.8	11	12	US-09-909-164-27	Sequence 27, Appl
30	41	78.8	11	12	US-09-909-164-61	Sequence 61, Appl
31	41	78.8	11	12	US-09-909-164-62	Sequence 62, Appl
32	40	76.9	11	12	US-09-909-164-21	Sequence 21, Appl
33	40	76.9	11	12	US-09-909-164-25	Sequence 25, Appl
34	40	76.9	11	12	US-09-909-164-31	Sequence 31, Appl
35	40	76.9	11	12	US-09-909-164-32	Sequence 32, Appl
36	40	76.9	11	12	US-09-909-164-35	Sequence 35, Appl
37	40	76.9	11	12	US-09-909-164-40	Sequence 40, Appl
38	40	76.9	11	12	US-09-909-164-41	Sequence 41, Appl
39	40	76.9	11	12	US-09-909-164-45	Sequence 45, Appl
40	40	76.9	11	12	US-09-909-164-46	Sequence 46, Appl
41	39	75.0	11	12	US-09-909-164-30	Sequence 30, Appl
42	39	75.0	11	12	US-09-909-164-34	Sequence 34, Appl
43	39	75.0	11	12	US-09-909-164-38	Sequence 38, Appl
44	39	75.0	11	12	US-09-909-164-39	Sequence 39, Appl
45	39	75.0	11	12	US-09-909-164-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-09-909-164-5
; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-5

Query Match 96.2%; Score 50; DB 12; Length 11;

OTHER INFORMATION: D-amino acids
JS-09-909-164-10

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 5

JS-09-909-164-47
; Sequence 47, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: valine-(CO)
JS-09-909-164-47

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 6

JS-09-909-164-48
; Sequence 48, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: leucine-(CO)
US-09-909-164-48

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 7

US-09-909-164-49
; Sequence 49, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norleucine-(CO)
US-09-909-164-49

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11

```
Db      1  EEVVPXGMSYS 11
|||||
RESULT 8
US-09-909-164-50
; Sequence 50, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: (s,s)-allothreonine-(CO)
US-09-909-164-51
Query Match          96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  EEVVPXGMSYS 11
|||||
Db      1  EEVVPXGMSYS 11
|||||

RESULT 10
US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: propynyl glycine-(CO)
US-09-909-164-52
Query Match          96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  EEVVPXGMSYS 11
|||||
Db      1  EEVVPXGMSYS 11
|||||

RESULT 11
US-09-909-164-8
; Sequence 8, Application US/09909164
; Publication No. US20020068702A1
```


RESULT 14

```
US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020086702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; FEATURE:
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-7
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EVVVPXGMYS 11
Db 1 EVVVPXGMYS 11
```

RESULT 15

```
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020086702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
```

```
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-11
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EVVVPXGMYS 11
Db 1 EVVVPXGMYS 11
```

```
Search completed: June 3, 2004, 12:57:15
Job time : 34.7333 secs
```


RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-877 <LYN>
A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A:Experimental source: strain 972h; cosmid c3H7
C:Genetics:
A:Gene: SPDB:SPBC3H7.02
A:Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| ||| :
Db 148 VVPQGMYSYA 156

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92189538; PMID:11546459
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 69.2%; Score 35; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
||| ||| :
Db 7 QVVPGINSY 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:g15026814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||| ||| :
Db 1276 EQKIPMGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TFP11) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of pistil-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MIL>
A:Cross-references: EMBL:U20592; NID:g924625; PIDN:AAA80497.1; PID:g924626
C:Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
||| ||| :
Db 32 DEVVPNGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-425 <WIL>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||| ||| :
Db 335 EQVIPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Mitcheimore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

A>Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
 A/Reference number: f58280; MUID:91187610; PMID:1901405
 A/Accession: S22293
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-670 <MIT>
 A/Cross-references: ENBL:X54250; NID:957519; PIDN:CAA38151.1; PID:957520
 A/Note: the authors did not translate the codon for residue 1
 C/Superfamily: HIV-BP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
 Best Local Similarity 66.7%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 ||| |||
 DB 376 VVPAGLTYS 384

RESULT 9
 f82691
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C/Accession: H82691
 C/Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A82515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: H82691
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-749 <SIM>
 A/Cross-references: GB:AE003967; GB:AE003849; NID:99106347; PIDN:AAF84162.1; GSPDB:GN001
 A/Experimental source: strain 9a5c
 A/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Doory, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A/Authors: Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 A.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 C/Map position: 15R
 C/Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase C

Query Match 67.3%; Score 35; DB 2; Length 749;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSY 10
 ||| |||
 DB 526 EVDPSGMSY 534

RESULT 10
 f34203
 A/Binding protein PRDII-BP1 - human
 A/Alternate names: major histocompatibility complex enhancer-binding protein 1
 C/Species: Homo sapiens (man)
 C/Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
 C/Accession: A34203; A34779
 C/Superfamily: Fan, C.M.; Maniatis, T.
 Genes Dev. 4, 29-42, 1990

A>Title: A DNA-binding protein containing two widely separated zinc finger motifs that i
 A/Reference number: A34203; MUID:90169514; PMID:2106471
 A/Accession: A34203
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2717 <FAN>
 A/Cross-references: ENBL:X51435; NID:938017; PIDN:CAA35798.1; PID:938018
 R/Baldwin Jr., A.S.; Leclair, K.P.; Singh, H.; Sharp, P.A.
 Mol. Cell. Biol. 10, 1406-1414, 1990
 A>Title: A large protein containing zinc finger domains binds to related sequence elemen
 A/Reference number: A34779; MUID:90205817; PMID:2108316
 A/Accession: A34779
 A/Status: preliminary; nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-16
 A/Cross-references: GB:N32019
 C/Superfamily: HIV-BP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 ||| |||
 DB 2405 VVPAGLTYS 2413

RESULT 11
 S54619
 A/hypothetical protein YOR013w - Yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
 C/Species: Saccharomyces cerevisiae
 C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 C/Accession: S54619; S66879
 R/de Haan, M.; Maarse, A.C.; Grivell, L.A.
 submitted to the EMBL Data Library, May 1995
 A/Reference number: S54617
 A/Accession: S54619
 A/Molecule type: DNA
 A/Residues: 1-156 <DEH>
 A/Cross-references: ENBL:X87331; NID:91041652; PIDN:CAA60762.1; PID:9829123
 R/de Haan, M.; Grivell, L.A.; Maarse, A.C.
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: S66877
 A/Accession: S66879
 A/Molecule type: DNA
 A/Residues: 1-156 <DEW>
 A/Cross-references: ENBL:Z74920; NID:91420109; PIDN:CAA99201.1; PID:91420111; MIPS:YOR01
 A/Experimental source: strain S288C
 C/Genetics:
 C/Cross-references: SGD:S0005539
 A/Map position: 15R
 C/Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
 Best Local Similarity 66.7%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSY 10
 ||| |||
 DB 50 EVNPLGNDY 58

RESULT 12
 H69491
 A/cell division inhibitor (mind-2) homolog - Archaeoglobus fulgidus
 C/Species: Archaeoglobus fulgidus
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 C/Accession: H69491
 R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, B.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

```

Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artisch, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woose, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: H69491
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-252 <KLE>
A;Cross-references: GB:AE000970; GB:AE000782; NID:G2689293; PIDN:AA889318.1; PID:G264860
C;Superfamily: cell division inhibitor mind

Query Match      65.4%; Score 34; DB 2; Length 252;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMS 9
Db 81 EVIPAGMS 88
|||:||||
|||:||||

RESULT 13
C82900
Probable ABC substrate-binding protein, iron UU359 [imported] - Ureaplasma urealyticum
C;Species: Ureaplasma urealyticum
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: C82900
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A;Reference number: A82870
A;Accession: C82900
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <GLA>
A;Cross-references: GB:AE002133; GB:AF222894; NID:G6899339; PIDN:AAF30768.1; GSPDB:GN001
A;Experimental source: serovar 3; biovar 1
C;Genetics:
A;Gene: ABCsbp-5; UU359
A;Genetic code: SGC3

Query Match      65.4%; Score 34; DB 2; Length 544;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGMSY 10
Db 135 EVVPHYLSY 144
|||||:|||
|||||:|||

RESULT 14
I40758
Hypothetical protein 1 - Campylobacter jejuni (fragment)
C;Species: Campylobacter jejuni
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40758; S47317
R;Hani, E.K.; Chan, V.L.
J. Bacteriol. 177, 2396-2402, 1995
A;Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohyd
A;Reference number: I40758; MUID:95247673; PMID:7730270
A;Accession: I40758
A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Molecule type: DNA
A;Residues: 1-94 <RES>
A;Cross-references: EMBL:Z36940; NID:G535805; PIDN:CAA85392.1; PID:G535806

Query Match      63.5%; Score 33; DB 2; Length 94;
Best Local Similarity 55.6%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 26 DIFPSGMSY 34
|||:||||
|||:||||

```

RESULT 15

```

E90544
50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: E90544
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, P.; Moszer, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: E90544
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-116 <KUR>
A;Cross-references: GB:AL445566; PID:gl4089674; PIDN:CAC13434.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2610
A;Genetic code: SGC3
C;Superfamily: Escherichia coli ribosomal protein L20

Query Match      63.5%; Score 33; DB 2; Length 116;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 68 VRPLGMSYS 76
|||:|||||
|||:|||||

```

Search completed: June 3, 2004, 11:59:58

Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	1058	1 CARB_FUSNN	Q8RG86 fusobacteri
2	37	71.2	877	1 SULH_SCHPO	O74377 schizosacch
3	36	69.2	102	1 Y1K_TYDVA	P31619 tobacco yel
4	36	69.2	1498	1 Y1A9_CLOAB	Q04351 clostridium
5	35	67.3	2717	1 ZEP1_HUMAN	P15922 homo sapien
6	34	65.4	788	1 CY14_NEUCR	P23622 neurospora
7	34	65.4	1499	1 A10A_HUMAN	O60312 homo sapien
8	33	63.5	116	1 RL20_MYCFU	Q98QV0 mycoplasma
9	33	63.5	165	1 Y349_ARCFU	O28330 archaeoglob
10	33	63.5	253	1 Y990_CAMJE	P45489 campylobact
11	33	63.5	280	1 CTX3_MOUSE	Q9d387 mus musculu
12	33	63.5	426	1 AROA_VIBCH	Q9ktb0 vibrio chol
13	33	63.5	466	1 CG37_SCHPO	O94740 schizosacch
14	33	63.5	478	1 GSR2_HUMAN	Q9nm55 homo sapien
15	33	63.5	890	1 BCN5_CLOPE	P08696 clostridium
16	33	63.5	1389	1 LTB5_MOUSE	Q8cg18 mus musculu
17	33	63.5	1394	1 LTB5_HUMAN	P22064 homo sapien
18	33	63.5	1401	1 RPOC_VIBCH	Q9kv29 vibrio chol
19	33	63.5	1595	1 LTB1_HUMAN	Q14766 homo sapien
20	33	63.5	1712	1 LTB1_RAT	Q00918 rattus norv
21	33	63.5	1713	1 LTB1_MOUSE	Q8cg19 mus musculu
22	32	61.5	288	1 CGD2_RAT	Q04827 rattus norv
23	32	61.5	289	1 CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	1 CGD2_MOUSE	P30280 mus musculu
25	32	61.5	291	1 CGD1_BRARE	Q90459 brachydanio
26	32	61.5	291	1 CGD1_XENLA	P50755 xenopus lae
27	32	61.5	291	1 CGD2_CHICK	P49706 gallus gall
28	32	61.5	291	1 CGD2_XENLA	P53782 xenopus lae
29	32	61.5	292	1 CGD1_CHICK	P55169 gallus gall
30	32	61.5	292	1 CGD3_HUMAN	P20281 homo sapien
31	32	61.5	295	1 CGD1_HUMAN	P24382 homo sapien
32	32	61.5	295	1 CGD1_MOUSE	P25322 mus musculu
33	32	61.5	295	1 CGD1_RAT	P39948 rattus norv

RESULT 1

ID	CARB_FUSNN	STANDARD;	PRT;	1058 AA.
AC	Q8RG86;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
DE	phosphate synthetase ammonia chain).			
GN	CARB OR FN0422.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;			
OC	Fusobacterium.			
OX	NCBI_TaxID=76856;			
EN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=ATCC 25586;			
RX	MEDLINE=21886394; PubMed=11889109;			
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Ponstein M., Kyripides N., Overbeek R.,			
RA	"Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."			
RT	J. Bacteriol. 184:2005-2018(2002).			
RL	-1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.			
CC	-1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	-1- PATHWAY: Arginine biosynthesis.			
CC	-1- PATHWAY: Pyrimidine biosynthesis; first step.			
CC	-1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).			
CC	-1- SIMILARITY: Belongs to the carb family.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AB010554; AAL94625.1; ALT_INT.			
DR	HAWAP; MF_01210; -; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; CPase_L.			
DR	InterPro; IPR005479; CPase_L_D2.			
DR	InterPro; IPR005480; CPase_L_D3.			
DR	InterPro; IPR005481; CPase_L_N.			
DR	InterPro; IPR004362; MCS_Like.			
DR	Pfam; PF00289; CPsase_L_chain; 2.			
DR	Pfam; PF02786; CPsase_L_D2; 2.			
DR	Pfam; PF02787; CPsase_L_D3; 1.			

ALIGNMENTS

34	32	61.5	427	1	TOLB_HAEIN
35	32	61.5	529	1	ENP3_HUMAN
36	32	61.5	691	1	S216_HUMAN
37	32	61.5	719	1	GSP_CRIFA
38	32	61.5	726	1	PRTP_HSV6U
39	32	61.5	726	1	PRTP_HSV6Z
40	32	61.5	759	1	SCY1_YEAST
41	32	61.5	920	1	EDD_RAT
42	32	61.5	993	1	VIA_TAV
43	32	61.5	1377	1	RHSA_ECOLI
44	32	61.5	1397	1	RHSC_ECOLI
45	32	61.5	1411	1	RHSE_ECOLI

P44677	haemophilus
O75355	homo sapien
Q9v616	homo sapien
P90518	crithidia f
P53284	human herpe
P53544	human herpe
P32784	saccharomyc
Q62671	rattus norv
P28931	tomato aspe
P16916	escherichia
P16918	escherichia
P16917	escherichia

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M81103; AAA47947.1; --
DR PIR; A42452; A42452.
DR InterPro; IPR002621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; 1.
DR Hypothetical protein.
DR SEQUENCE 102 AA; 11178 MW; A408CFIE0AF5B867 CRG64;

Query Match 69.2%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 1.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Y 2 EVVVPXGMSYS 11
b 7 QVPSGINS 16
: ||| |::||

RESULT 4
1A9 CLOAB STANDARD; PRT; 1498 AA.
C Q04351;
D YIA9 CLOAB STANDARD; PRT; 1498 AA.
T 01-FEB-1994 (Rel. 28, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
B Hypothetical protein CAC3709.
N CAC3709.
S Clostridium acetobutylicum.
X Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
X Clostridium.
X NCBI_TaxID=1488;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
X MEDLINE=21359325; PubMed=11466286;
A Noelling J., Breston G., Omeichenko M.V., Makarova K.S., Zeng Q.,
A Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
A Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
A Bennett G.N., Koonin E.V., Smith D.R.;
T "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.";
J. Bacteriol. 183:4823-4838(2001).
N [2]
P SEQUENCE OF 1-108 FROM N.A.
C STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
X MEDLINE=93273706; PubMed=8501044;
A Sauer U., Duerre P.;
T "Sequence and molecular characterization of a DNA region encoding a small heat shock protein of Clostridium acetobutylicum.";
J. Bacteriol. 175:3394-3400(1993).
L -1- SIMILARITY: Contains 2 PfsK domains.
C -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts in positions 76 and 106.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AE007866; AKS1629.1; --
DR EMBL; X65276; CAA46379.1; ALT_FRAME.
DR PIR; B97355; B97355.
DR InterPro; IPR002543; PfsK_SpoIIIE.

DR Pfam; PF01580; PfsK_SpoIIIE; 2.
DR PROSITE; PS0901; FTSK; 2.
KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
FT DOMAIN 655 857 FTSK 1.
FT DOMAIN 1001 1188 FTSK 2.
FT NP_BIND 675 692 ATP (POTENTIAL).
SQ SEQUENCE 1498 AA; 169968 MW; FF42037A335A9649 CRG64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Q 1 EVVVPXGMSY 10
b 1276 EQIPMGMSY 1285
: : ||| |

RESULT 5
ZEPI HUMAN STANDARD; PRT; 2717 AA.
ID ZEPI HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE [PRDII-BP1]
DE [PRDII-BP1]
GN HIVP1 OR ZNF40.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90169514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K., Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142.
RX MEDLINE=92232694; PubMed=1567844;
RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E., Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992). THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE 5'-GGGACTTCC-3', WHICH IS FOUND IN THE ENHANCER ELEMENTS OF NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1. IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: By mitogens and phorbol ester.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVP2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - between the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X51435; CAA35798.1; --
 PIR; A34203; A34203.
 PDB; 3ZNF; 15-JAN-92.
 PDB; 4ZNF; 15-JAN-92.
 PDB; 1BBO; 31-OCT-93.
 TRANSFAC; T00497; --
 Genew; HGNC:4920; HIVEP1.
 MIM; 194540; --
 GO; GO:0005634; C:nucleus; TAS.
 GO; GO:0003677; F:DNA binding; TAS.
 InterPro; IPR007087; Znf_C2H2.
 Pfam; PF00096; zf-C2H2; 5.
 SMART; SM00355; Znf_C2H2; 4.
 PROSITE; PS0028; ZINC_FINGER_C2H2_1; 4.
 PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 Nuclear protein; Repeat; 3D-structure.
 ZN_FING 406 428
 ZN_FING 434 456
 ZN_FING 958 981
 ZN_FING 2087 2109
 ZN_FING 2115 2139
 ZN_FING 2139 2139
 DOMAIN 803 806
 STRAND 2088 2088
 TURN 2090 2092
 STRAND 2095 2095
 HELIX 2099 2108
 TURN 2109 2109
 STRAND 2115 2116
 STRAND 2123 2124
 HELIX 2127 2135
 SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

DB 2405 VVPAGLTYS 2413

RESULT 6

CV14_NEUCR
 ID CV14_NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate permease II.
 GN CVS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 CX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=91129256; PubMed=1825178;
 RA Katter J.S., Jarai G., Fu Y.-H., Marzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 elements of *cyS-14*, the structural gene for sulfate permease II in
Neurospora crassa,"
 RL Biochemistry 30:1780-1787(1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RP MEDLINE=94188926; PubMed=8140616;
 RA Sandel N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, *Neurospora crassa* sulphate

permease II and a putative human tumour suppressor.";
 Trends Biochem. Sci. 19:19-19(1994).
 CC FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC SUBCELLULAR LOCATION: Integral membrane protein.
 CC INDUCTION: Highly expressed, but only in cells subject to sulfur
 limitation, and it is turned on by the positive-acting Cys-3
 sulfur regulatory protein.
 CC MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC SIMILARITY: Belongs to the SLC26A/Sulph transporter (TC 2.A.53)
 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M59167; AAA33615.1; ALT_SEQ.

PIR; A37956; A37956.

InterPro; IPR001302; Sulph transport.

Pfam; PF00916; Sulfate_transp; 1.

TIGRfams; TIGR00815; sulp; 1.

PROSITE; PS01130; SLC26A; 1.

Transport; Transmembrane; Glycoprotein.

TRANSMEM 71 91 POTENTIAL.

TRANSMEM 103 123 POTENTIAL.

TRANSMEM 128 148 POTENTIAL.

TRANSMEM 171 191 POTENTIAL.

TRANSMEM 193 213 POTENTIAL.

TRANSMEM 271 291 POTENTIAL.

TRANSMEM 326 346 POTENTIAL.

TRANSMEM 363 383 POTENTIAL.

TRANSMEM 451 471 POTENTIAL.

TRANSMEM 474 494 POTENTIAL.

CARBOHYD 23 23 N-LINKED (GLCNAC...) (POTENTIAL).

CARBOHYD 578 578 N-LINKED (GLCNAC...) (POTENTIAL).

SEQUENCE 789 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

DB 90 VVPQGMAYA 98

RESULT 7

AL0A_HUMAN
 ID AL0A_HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; Q96914;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
 DE (Aminophospholipid translocase VA)
 GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=21225279; PubMed=11326269;
 RA Meguro M., Kashiwagi A., Mitsuoka K., Nakao M., Kondo I., Saitoh S.,
 Oshimura M.;
 RT "A novel maternally expressed gene, ATP10C, encodes a putative
 aminophospholipid translocase associated with Angelman syndrome.";
 RL Nat. Genet. 28:19-20(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

MDLINE=21313119; PubMed=11353404;
RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
RT "The human aminophospholipid-translocating ATPase gene ATP10C maps
RL adjacent to UBE3A and exhibits similar imprinted expression.";
RN Am. J. Hum. Genet. 68:1501-1505 (2001).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lottolano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grilowinski J., Schmutz J., Myers R.N.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
RN SEQUENCE OF 337-1499 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RD Nucleic Acids Res. 26:1531-1539 (1998).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- TISSUE SPECIFICITY: Widely expressed, with highest levels in
CC kidney, followed by lung, brain, prostate, testis, ovary and
CC small intestine.
CC -!- DISEASE: Defects in ATP10A are a cause of Angelman syndrome (AS)
CC [MIM:105830]; also known as 'happy puppet syndrome'. AS is
CC characterized by features of severe motor and intellectual
CC retardation, microcephaly, ataxia, frequent jerky limb movements
CC and flapping of the arms and hands, hypotonia, hyperactivity,
CC hypopigmentation, seizures, absence of speech, frequent smiling
CC and episodes of paroxysmal laughter, and an unusual facies
CC characterized by macrostomia, a large mandible and open-mouthed
CC expression, a great propensity for protruding the tongue ('tongue
CC thrusting'), and an occipital groove.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB051338; BAB7392.1; -;
CC EMBL; AY029504; AAK33100.1; -;
CC EMBL; AY029487; AAK33100.1; JOINED.
CC EMBL; AY029488; AAK33100.1; JOINED.
CC EMBL; AY029489; AAK33100.1; JOINED.
CC EMBL; AY029490; AAK33100.1; JOINED.
CC EMBL; AY029491; AAK33100.1; JOINED.
CC EMBL; AY029492; AAK33100.1; JOINED.
CC
CC DR EMBL; AY029493; AAK33100.1; JOINED.
CC DR EMBL; AY029494; AAK33100.1; JOINED.
CC DR EMBL; AY029495; AAK33100.1; JOINED.
CC DR EMBL; AY029496; AAK33100.1; JOINED.
CC DR EMBL; AY029497; AAK33100.1; JOINED.
CC DR EMBL; AY029498; AAK33100.1; JOINED.
CC DR EMBL; AY029499; AAK33100.1; JOINED.
CC DR EMBL; AY029500; AAK33100.1; JOINED.
CC DR EMBL; AY029501; AAK33100.1; JOINED.
CC DR EMBL; AY029502; AAK33100.1; JOINED.
CC DR EMBL; AY029503; AAK33100.1; JOINED.
CC DR EMBL; BC052251; AAK33100.1; JOINED.
CC DR EMBL; AB011138; BAA25492.1; -;
CC DR Genew; HGNC:13542; ATP10A.
CC DR MIM; 605855; -;
CC DR MIM; 105830; -;
CC DR GO; GO:0016021; C: integral to membrane; NAS.
CC DR GO; GO:0004012; F: phospholipid-translocating ATPase activity; NAS.
CC DR GO; GO:0008360; P: regulation of cell shape; NAS.
CC DR InterPro; IPR001757; ATPase_E1_E2.
CC DR InterPro; IPR006539; Flippase.
CC DR Pfam; PF00702; Hydrolase; 1.
CC DR PRINTS; PR00119; CATAPPAE.
CC DR TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
CC DR TIGRFAMs; TIGR01494; ATPase P-type; 6.
CC DR PROSITE; PS00154; ATPase_E1_E2; 1.
CC KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
CC MW Multigene family.
CC FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 87 106 POTENTIAL.
CC FT DOMAIN 107 110 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 111 128 POTENTIAL.
CC FT DOMAIN 129 309 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 310 332 POTENTIAL.
CC FT DOMAIN 333 362 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 363 384 POTENTIAL.
CC FT DOMAIN 385 1087 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 1088 1108 POTENTIAL.
CC FT DOMAIN 1109 1119 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1120 1140 POTENTIAL.
CC FT DOMAIN 1141 1170 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 1171 1192 POTENTIAL.
CC FT DOMAIN 1193 1199 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1200 1222 POTENTIAL.
CC FT DOMAIN 1223 1228 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 1229 1249 POTENTIAL.
CC FT DOMAIN 1250 1267 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1268 1292 POTENTIAL.
CC FT DOMAIN 1293 1499 CYTOPLASMIC (POTENTIAL).
CC FT MOD RES 427 427 PHOSPHORYLATION (BY SIMILARITY).
CC FT METAL 1031 1031 MAGNESIUM (BY SIMILARITY).
CC FT METAL 1035 1035 MAGNESIUM (BY SIMILARITY).
CC FT DOMAIN 467 470 POLY-GLU.
CC FT CONFLICT 388 388 Q -> R (IN REF. 4).
CC FT SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
CC
CC Query Match 65.4%; Score 34; DB 1; Length 1499;
CC Best Local Similarity 72.7%; Pred. No. 70;
CC Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 1 EEVVPXGMSYS 11
CC |||||
CC DB 469 EEVPRGGSVS 479
CC
CC RESULT 8
CC RL20 MYCPU
CC ID RL20 MYCPU STANDARD; PRT; 116 AA.
CC AC Q98QV0;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 50S ribosomal protein L20.
GN RPLT OR MYPU 2610.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Gallisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.; genome sequence of the murine respiratory pathogen
RT "the complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -/- FUNCTION: This protein binds directly to 23S ribosomal RNA and is
CC necessary for the in vitro assembly process of the 50S ribosomal
CC subunit. It is not involved in the protein synthesizing functions
CC of that subunit (By similarity).
CC -/- SIMILARITY: Belongs to the L20P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL445563; CAC13434.1; -
DR PIR; E90544; E90544.
DR MyDulist; MYPU 2610; -
DR HAMAP; MF_00382; -; 1
DR InterPro; IPR005843; Ribosomal L20.
DR InterPro; IPR005842; Ribosomal L20b/o.
DR Pfam; PF00453; Ribosomal L20; 1.
DR PRINTS; PR00062; RIBOSOMAL20.
DR ProDom; PD002389; L20; 1.
DR TIGRfam; TIGR01032; rplL bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18f14 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 8.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPXGMSYS 11
Db 68 VRFLGMSYS 76
RESULT 9
YJ49 ARCFU
ID YJ49_ARCFU STANDARD; PRT; 165 AA.
AC 028330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN Ar1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Karlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.F., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000968; AAB89307.1; -
DR PIR; D69493; D69493.
DR TIGR; AF1949; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 165 AA; 17588 MW; B6C17054810ADB8 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 REVPXGMSY 10
Db 60 EESIPDGASY 69
RESULT 10
Y990 CAMJE
ID Y990_CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990c.
GN Cj0990c.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Peltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream A.M., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli.";
RN J. Bacteriol. 177:2396-2402(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AL139076; CAB73246.1; --
 DR EMBL; Z36940; CAA85392.1; --
 DR PIR; C81374; C81374.
 DR PIR; C81374; C81374.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 29783 MW; F95D3FF265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
 Best Local Similarity 55.6%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

2y 2 EVVPXGMSY 10
 :|||
 185 DIFPSGMSY 193

RESULT 11

CTX3 MOUSE

ID CTX3 MOUSE STANDARD; PRT; 280 AA.
 JT 28-FEB-2003 (Rel. 41, Created)
 JT 28-FEB-2003 (Rel. 41, Last sequence update)
 JT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 EN C20ORF103.
 DS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 XX NCBI_TaxID=10090;
 XN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=20185660; PubMed=11217851;
 RA Kawai J. Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gijobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiava H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staali F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 XN [2]

SEQUENCE FROM N.A. (ISOFORM 2).

TISSUE=Mammary fibroblast;
 MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -|- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AK014127; BAB29169.1; --
 DR EMBL; AK018222; BAB31124.1; ALT_FRAME.
 DR EMBL; BC004791; AAB04791.1; --
 DR MGI; MGI:1920368; 311003SN03Rik.
 DR MGI; MGI:1923411; 6330527006Rik.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 280 PROTEIN C20ORF103 HOMOLOG.
 FT DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 236 256 POTENTIAL.
 FT DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPPLIC 1 118 Missing (in isoform 2).
 FT /FTID=VSP_003820.
 FT CONFLICT 221 221 E -> V (IN REF. 1; BAB311124).
 FT CONFLICT 230 230 Q -> P (IN REF. 1; BAB311124).
 FT CONFLICT 238 238 P -> A (IN REF. 1; BAB311124).
 SQ SEQUENCE 280 AA; 31721 MW; FA11D7BF9DF5CCEFC CRC64;

Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMSY 10
 :|||
 Db 173 VTPAGMSY 180

RESULT 12

AROA_VIBCH

ID AROA_VIBCH STANDARD; PRT; 426 AA.
 AC Q9KR80;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.15) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR VC1732.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=E1 TOR N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483 (2000).
CC -1- CATABOLIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004251; AAF94882.1; -.
CC DR PIR; D82163; D82163.
CC DR TIGR; VC1732; -.
CC DR HAVAP; MF 00210; -.
CC DR InterPro; IPR006264; AroA.
CC DR InterPro; IPR001986; EPSP synth.
CC DR Pfam; PF00275; EPSP synthase; 1.
CC DR PRODOM; PD001867; EPSP synthase; 1.
CC DR TIGRFAMs; TIGR01356; AroA; 1.
CC DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
CC DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
CC DR Aromatic amino acid biosynthesis; Transferase; Complete proteome.
KW AROMATIC AMINO ACID BIOSYNTHESIS; TRANSFERASE; COMPLETE PROTEOME.
SQ SEQUENCE 426 AA; 46101 MW; 38852D8483BFELC3 CRC64;
-----
Query Match 63.5%; Score 33; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGMSY 10
DB 223 EFVIPAGQSY 232
-----
RESULT 13
CC37 SCHPO STANDARD; PRT; 466 AA.
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE subunit) (Cell division control protein 37).
GN CDC37 OR SPBC9B6.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Westwood P.K., Preston N.C., Fantes P.A.;
RT "Schizosaccharomycetes pombe cdc37 gene.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Haidvo J., Hodgson G.,
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James L., Jones M., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren I., Whitehead S.,
RA Woodward J., Wolckaert G., Aert R., Robben J., Grymonprez B.,
RA Mewaldt J., Vansteelandt E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Burnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
-----
[3]
RN RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=22745456; PubMed=12861001;
RA Tatebe H., Shiozaki K.;
RT Identification of cdc37 as a novel regulator of the stress-responsive
RT mitogen-activated protein kinase.";
RL Mol. Cell. Biol. 23:5132-5142 (2003).
CC -1- FUNCTION: Co-chaperone that binds to numerous kinases and promotes
CC their interaction with the Hsp90 complex, resulting in
CC stabilization and promotion of their activity.
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with sty1.
CC -1- SUBCELLULAR LOCATION: Nuclear, and cytoplasmic. When in the
CC nucleus associated with chromatin.
CC -1- SIMILARITY: Belongs to the CDC37 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ132377; CAB38758.1; -.
CC DR EMBL; AJ132376; CAB38757.1; -.
CC DR EMBL; AL049769; CAB42371.2; -.
CC DR PIR; T43653; T43653.
CC DR GeneDB; Spombe; SPBC9B6.10; -.
CC DR InterPro; IPR004918; Cdc37.
CC DR Pfam; PF03234; Cdc37; 1.
CC DR Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
CC subunit) (Cell division control protein 37).
KW CHAPERONE; CELL DIVISION; CELL CYCLE; NUCLEAR PROTEIN.
SQ SEQUENCE 466 AA; 52554 MW; 647238B34CABB3C5 CRC64;
-----
Query Match 63.5%; Score 33; DB 1; Length 466;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGMSY 10
DB 98 DSAIFGMSY 107
-----
RESULT 14

```

SSR2_HUMAN
ID GSR2_HUMAN STANDARD; PRT: 478 AA.
AC Q9NZM5; Q9HAX6; Q9NP21; Q9NP4; Q9UF12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
EN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.,
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
region.";
RL Genomics 64:44-50(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
RX MEDLINE=99214316; PubMed=10196275;
RA Bruni R., Fineschi B., Ogle W.O., Roizman B.;
RT "A novel cellular protein, p60, interacting with both herpes simplex
virus 1 regulatory proteins ICP22 and ICP0 is modified in a
cell-type-specific manner and is recruited to the nucleus after
infection.";
RL J. Virol. 73:3810-3817(1999).
RN [4]
RP SEQUENCE OF 12-478 FROM N.A.
RA Andreu N., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 218-477 FROM N.A.
RC TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in heart and
pancreas, moderate levels in placenta, liver, skeletal muscle, and
kidney, and low levels in brain and lung.
CC -!- SIMILARITY: Belongs to the GLTSCR2 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF182076; AAF62873.1; -
DR EMBL; BC004229; AAH04229.1; -
DR EMBL; BC006311; AAH06311.1; -
DR EMBL; BC010095; AAH10095.1; -
DR EMBL; AF296124; AAG30413.1; -
DR EMBL; AF296124; AAG30413.1; -
DR EMBL; AL359335; CAB94786.1; -
DR EMBL; AL359336; CAB94787.1; -
DR EMBL; AL122063; CAB59242.1; -
DR SWISS-2DPAGE; Q9NZM5; HUMAN.
DR Genew; HGNC:4333; GLTSCR2.
DR MIM; 605691; -
DR GO; GO:0005622; C:intracellular; NAS.
KW Nuclear protein; Polymorphism.
FT VARIANT 389 389
FT CONFLICT 4 6 /FTID=VAR 011486.
FT CONFLICT 9 9 GGS -> HEG (IN REF. 2; AAH04229).
FT CONFLICT 146 191 G -> R (IN REF. 3).
FT CONFLICT 198 215 RRKEQLWEKLAKOGELPREVREAOALNPSATRAKPGPOD
TTVERP -> SGRSSYGRSWPSRASSPGGAGQSPVAQPCFN
KGNPAPGHRKIA (IN REF. 3).
FT CONFLICT 235 235 A -> S (IN REF. 2; AAH04229).
FT CONFLICT 417 417 G -> H (IN REF. 3).
FT CONFLICT 433 477 PEGNIILDRKSFQRRNMIEPRERAKFKKVKVGLVEKRAF
REIQ -> VLTVSCRGAPCPVMTPSLLPVPRGYGRHHCPC
WAGFVGPMERG (IN REF. 5).
FT CONFLICT 434 478 EGNILDRFKSFQRRNMIEPRERAKFKKVKVGLVEKRAF
ETQL -> RQHSFETGSAFRGGI (IN REF. 3).
SQ SEQUENCE 478 AA; 54417 MW; 7F18923E348CB52B CRC64;
Query Match 63.5%; Score 33; DB 1; Length 478;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Oy 2 EVVPXGMSYS 11
Db 239 EVAPAGASYN 248

RESULT 15
BCN5_CLOPE STANDARD; PRT: 890 AA.
AC P08696;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Bacteriocin BCN5.
GN BGN.
OS Clostridium perfringens.
OG Plasmid pIP404.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=88336297; PubMed=2901768;
RA Garnier T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garnier T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium

```
RT perfringens and molecular genetic analysis of the
RL bacteriocin-encoding gene."
RN J. Bacteriol. 168:1189-1196(1986).
RP [3]
RC SEQUENCE OF 1-14 FROM N.A.
RX STRAIN=CPN50;
RA MEDLINE=89039249; PubMed=2460717;
RT Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro."
RL Mol. Microbiol. 2:607-614(1988).
CC -!- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -!- INDUCTION: By UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -
DR EMBL; M32882; AAA98249.1; -
DR PIR; A30481; A30481
DR InterPro; IPR000834; Peptidase_M14.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00287; SH3b; 3.
DR Antibiotic; Bacteriocin; Plasmid.
FT DOMAIN 815 869 HYDROPHOBIC.
SQ SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred.No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 170 EVVPGGFTY 178
|||||:|
```

Search completed: June 3, 2004, 11:49:50
Job time : 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06; Search time 29.8667 Seconds

(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52

Sequence: 1 EVVFXGMSYS 11

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	75.0	1044	16	Q8DIH0	Q8DIH0 synechococc
2	38	73.1	344	16	Q815A7	Q815A7 bacillus ce
3	38	73.1	387	16	Q98FX1	Q98FX1 rhizobium 1
4	38	73.1	3472	1	Q74056	Q74056 cenarchaeum
5	37	71.2	840	3	Q9UR18	Q9UR18 schizosacch
6	37	71.2	1123	16	Q8EWD4	Q8EWD4 mycoplasma
7	36	69.2	471	11	Q8R126	Q8R126 mus musculu
8	36	69.2	484	11	Q8VD18	Q8VD18 mus musculu
9	36	69.2	484	11	Q8BTX4	Q8BTX4 mus musculu
10	36	69.2	484	11	Q8BK35	Q8BK35 mus musculu
11	36	69.2	559	16	Q83T9	Q83T9 enterococcu
12	36	69.2	1399	16	Q889X7	Q889X7 pseudomonas
13	35	67.3	225	10	Q40129	Q40129 lycopersico
14	35	67.3	245	16	Q7V6Q4	Q7V6Q4 prochloroco
15	35	67.3	425	5	Q5XVK4	Q5XVK4 caenorhabd
16	35	67.3	495	11	Q8C1D7	Q8C1D7 mus musculu

17	35	67.3	555	4	Q7Z6R0	Q7Z6R0 homo sapien
18	35	67.3	583	5	Q9BH83	Q9BH83 plasmodium
19	35	67.3	583	5	Q9BHA5	Q9BHA5 plasmodium
20	35	67.3	583	5	Q815S7	Q815S7 plasmodium
21	35	67.3	670	11	Q01487	Q01487 rattus ratt
22	35	67.3	747	16	Q8PM16	Q8PM16 xanthomonas
23	35	67.3	747	16	Q8PAT2	Q8PAT2 xanthomonas
24	35	67.3	749	16	Q9PDM6	Q9PDM6 xyella fas
25	35	67.3	1902	4	Q14122	Q14122 homo sapien
26	34	65.4	156	3	Q12479	Q12479 saccharomyc
27	34	65.4	219	17	Q971S2	Q971S2 sulfolobus
28	34	65.4	252	17	Q28342	Q28342 archaeoglob
29	34	65.4	290	4	Q96MD1	Q96MD1 homo sapien
30	34	65.4	387	16	Q92MD6	Q92MD6 rhizobium m
31	34	65.4	489	4	Q81YM3	Q81YM3 homo sapien
32	34	65.4	541	16	Q98BP5	Q98BP5 rhizobium 1
33	34	65.4	544	16	Q9PQD2	Q9PQD2 ureaplasma
34	34	65.4	731	16	Q7UWU7	Q7UWU7 rhodopirell
35	34	65.4	842	3	Q9URR4	Q9URR4 penicillium
36	34	65.4	899	16	Q8G415	Q8G415 bifidobacte
37	34	65.4	1049	16	Q8XT05	Q8XT05 raietonia s
38	34	65.4	1400	16	Q87KQ5	Q87KQ5 vibrio para
39	33	63.5	143	17	Q8TX62	Q8TX62 methanopyru
40	33	63.5	166	16	Q8PPP5	Q8PPP5 xanthomonas
41	33	63.5	193	2	Q8VUA8	Q8VUA8 lactococcus
42	33	63.5	208	2	Q8KTO4	Q8KTO4 candidatus
43	33	63.5	209	16	Q8RE56	Q8RE56 fusobacteri
44	33	63.5	251	13	Q7SY67	Q7SY67 xenopus lae
45	33	63.5	282	16	Q7U552	Q7U552 synechococc

ALIGNMENTS

RESULT 1

Q8DIH0 PRELIMINARY; PRT; 1044 AA.

AC Q8DIH0; 01-MAR-2003 (TREMREL. 23, Created)

DT 01-MAR-2003 (TREMREL. 23, Last sequence update)

DT 01-JUN-2003 (TREMREL. 24, Last annotation update)

DE Multidrug efflux transporter.

GN TLL1618.

OS Synechococcus elongatus (Thermosynechococcus elongatus).

OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.

OX NCBI_TaxID=32046;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BP-1;

EX MEDLINE=2225144; PubMed=12240834;

RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,

Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,

Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,

Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the thermophilic cyanobacterium

Thermosynechococcus elongatus BP-1.";

RL DNA Res. 9:123-130(2002).

DR EMBL; AP005374; BAC09170.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005215; P:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR001036; Acriflavin_res.

DR InterPro; IPR004764; HAE1.

DR Pfam; PF00873; ACR_tran; 1.

DR PRINTS; PR00702; ACRIFLAVINRP.

DR TIGRFAMs; TIGR00915; 2A0602; 1.

KW Complete proteome.

SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 75.0%; Score 39; DB 16; Length 1044;

Best Local Similarity 63.6%; Pred. No. 28;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 EVVVPXGMSYS 11
DB 843 EAVLPNGIGYS 853
|||:|:|
|||:|:|

RESULT 2
Q815A7 Q815A7 PRELIMINARY; PRT; 344 AA.
AC Q815A7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BC5259.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017015; AAP12123.1; -.
DR InterPro; IPR00437; PROKALIPROTEIN.S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 344 AA; 38539 MW; C55268ACB7225995 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 344;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
DB 152 EETAPLGLSY 161
|||:|:|
|||:|:|

RESULT 3
Q98FX1 Q98FX1 PRELIMINARY; PRT; 387 AA.
AC Q98FX1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hippurate hydrolase.
GN MLR3583.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50445.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; peptidase_M20;

Pfam; PF01546; Peptidase_M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 387 AA; 41190 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
DB 367 DEAIPIHGMSY 376
|||:|:|
|||:|:|

RESULT 4
O74056 O74056 PRELIMINARY; PRT; 3472 AA.
AC O74056;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
OC Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=B;
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL; AF083072; AAC62699.1; -.
DR FIR; T31308; T31308.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0008810; F:transport; IEA.
DR InterPro; IPR000515; BPD_transp.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Hypothetical protein.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
DB 2294 EDVIRGISFS 2304
|||:|:|
|||:|:|

RESULT 5
Q9URY8 Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=972h-;
RC Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;

```


IL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
R EMBL; AL132779; CAB60015.1; -
R PIR; T39116; T39116.
R GeneDB SPombe; SPAC869.05c; -
R GO; GO:0016020; C:membrane; IEA.
R GO; GO:0008271; F:sulfate porter activity; IEA.
R GO; GO:0008272; P:sulfate transport; IEA.
R InterPro; IPR002645; STAS.
R InterPro; IPR001902; Sulph_transpt.
R Pfam; PF01740; STAS; 1.
R TIGRFAMs; TIGR00815; sulP; 1.
R PROSITE; PS0801; STAS; 1.
Q SEQUENCE 840 AA; 93517 MW; ED4933E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 3 VWPXGMSYS 11
||| |||
b 135 VWPQMSYA 143

RESULT 6
Q8EWD4 PRELIMINARY; PRT; 1123 AA.
D Q8EWD4;
C Q8EWD4;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E MYPE 2560 paralogs, 57%
N MYPE2710.
S Mycoplasma penetrans.
C Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
X NCBI_TaxID=28227;
X [1]
P SEQUENCE FROM N.A.
R STRAIN=HF-2;
X MEDLINE=22354719; PubMed=12466555;
A Sasaki Y, Ishikawa J., Yamashita A., Oshima K., Kenri T., Furiya K.,
A Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
T "The complete genomic sequence of Mycoplasma penetrans, an
intracellular bacterial pathogen in humans."
U Nucleic Acids Res. 30:5293-5300(2002).
R EMBL; AP004171; BAC44062.1; -
R InterPro; IPR008985; ConA like lec.gl.
R InterPro; IPR007326; Lipoprotein_17.
R Pfam; PF04200; Lipoprotein_17; 3.
C Complete proteome.
Q SEQUENCE 1123 AA; 123636 MW; A4D707330E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
Best Local Similarity 70.0%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Y 2 EWPXGMSYS 11
||| |||
b 658 EWPVGLSYS 667

RESULT 7
Q8R126 PRELIMINARY; PRT; 471 AA.
D Q8R126;
C Q8R126;
T 01-JUN-2002 (TrEMBLrel. 21, Created)
T 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Hypothetical protein (Fragment).
N GLTSCR2.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
D Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;

OX NCBI_TaxID=10090;
R [1]
R SEQUENCE FROM N.A.
R TISSUE=Liver;
R Strausberg R.;
R Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; BC025810; AAH25810.1; -
R MGD; MGI:2154441; GLTSCR2.
R Hypothetical protein.
R NON TER 1
Q SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Q 2 EVVPXGMSYS 11
||| |||
b 226 EVIPAGASYN 235

RESULT 8
Q8VD18 PRELIMINARY; PRT; 484 AA.
D Q8VD18;
C Q8VD18;
T 01-MAR-2002 (TrEMBLrel. 20, Created)
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Similar to glioma tumor suppressor candidate region gene 2.
N GLTSCR2 OR AW536441.
C Mus musculus (Mouse).
O Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
R [1]
R SEQUENCE FROM N.A.
R TISSUE=Salivary gland;
R Strausberg R.;
R Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; BC017637; AAH17637.1; -
R MGD; MGI:2154441; GLTSCR2.
Q SEQUENCE 484 AA; 55835 MW; BB945F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Q 2 EVVPXGMSYS 11
||| |||
b 239 EVIPAGASYN 248

RESULT 9
Q8BTX4 PRELIMINARY; PRT; 484 AA.
D Q8BTX4;
C Q8BTX4;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Similar to glioma tumor suppressor CANDIDATE region gene 2
protein.
N GLTSCR2.
C Mus musculus (Mouse).
O Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;

RN [1]
R SEQUENCE FROM N.A.
R STRAIN=NOD; TISSUE=Thymus;
R MEDLINE=22354683; PubMed=12466851;
R THE FANTOM Consortium;
R the RIKEN Genome Exploration Research Group Phase 1 & 11 Team;
R "Analysis of the mouse transcriptome based on functional annotation of

```

RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK089461; BAC40367.1; -.
DR MGD; MGI:2154441; GLScr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EECAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 10
Q8BK35 PRELIMINARY; PRT; 484 AA.
ID Q8BK35
AC Q8BK35
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
DE GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK077341; BAC36760.1; -.
DR MGD; MGI:2154441; GLScr2.
SQ SEQUENCE 484 AA; 55792 MW; 5B67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 11
Q839T9 PRELIMINARY; PRT; 559 AA.
ID Q839T9
AC Q839T9
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Pheromone binding protein, putative.
DS EF0063.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;

```

```

RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis."
RL Science 299:2071-2074 (2003).
DR EMBL; AE016947; AAO79943.1; -.
DR TIGR; EF0063.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 559 AA; 61476 MW; CC15418D33D53DE7 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 559;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 351 LIPEGMSYS 359

RESULT 12
Q889X7 PRELIMINARY; PRT; 1399 AA.
ID Q889X7
AC Q889X7
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE DNA-directed RNA polymerase, beta' subunit.
DE RPOC OR PSPT00620.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buehl R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RA Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016858; AAO34162.1; -.
DR TIGR; PSPT00620; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
KW DNA-directed RNA polymerase; Complete proteome.
SQ SEQUENCE 1399 AA; 154733 MW; 26178BD653102ADE CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1399;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 10
Db 581 QVVPAGLSY 589

```

```
RESULT 13
240129 ID Q40129 PRELIMINARY; PRT; 225 AA.
AC Q40129;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein precursor.
DS Lycopersicon esculentum (tomato).
DC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
DC Lamiales; Solanales; Solanaceae; Solanum.
XX NCBI_TaxID=4081;
XX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=VF36; TISSUE=Pistil;
RX MEDLINE=95375233; PubMed=7647301;
RA Milligan S.B., Gasser C.S.;
RT "Nature and regulation of pistil-expressed genes in tomato.";
RL Plant Mol. Biol. 28:691-711(1995).
DR EMBL; U20592; AAA80497.1; -.
DR PIR; S57810; S57810.
DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINBTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; ST1.1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KW Hypothetical protein; signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

2y 1 EEVVPXGMSY 11
:|||||:
Db 32 DEVPVNGTKYA 42

RESULT 14
Q7V6Q4 ID Q7V6Q4 PRELIMINARY; PRT; 245 AA.
AC Q7V6Q4;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phospholipid and glycerol acyltransferase (From 'motifs_6.msff').
GN PM11092.
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Allgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.P., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
DR EMBL; BX52098; CAE21267.1; -.
KW Acyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 245 AA; 26907 MW; 106F7C4CBE2C6427 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 245;
```

```
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EEVVPXGMSY 11
:|||||:
Db 179 QVVPVGLGYS 188

RESULT 15
Q9XVK4 ID Q9XVK4 PRELIMINARY; PRT; 425 AA.
AC Q9XVK4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
OX [1]
RN SEQUENCE FROM N.A.
RP Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81109; CAB03241.1; -.
DR PIR; T24111; T24111.
DR WormPep; R10D12.10; CE12690.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSY 10
:|||||:
Db 335 EQIVPGLQY 344

Search completed: June 3, 2004, 11:57:30
Job time : 29.8667 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-7
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	96.4	11	5 ABB80523	Abb80523 Hepatitis
2	54	96.4	11	5 ABB80558	Abb80558 Hepatitis
3	54	96.4	11	5 ABB80560	Abb80560 Hepatitis
4	54	96.4	11	5 ABB80527	Abb80527 Hepatitis
5	49	87.5	11	5 ABB80537	Abb80537 Hepatitis
6	49	87.5	11	5 ABB80541	Abb80541 Hepatitis
7	48	85.7	11	5 ABB80546	Abb80546 Hepatitis
8	48	85.7	11	5 ABB80554	Abb80554 Hepatitis
9	48	85.7	11	5 ABB80550	Abb80550 Hepatitis
10	48	85.7	11	5 ABB80555	Abb80555 Hepatitis
11	46	82.1	11	5 ABB80532	Abb80532 Hepatitis
12	46	82.1	11	5 ABB80531	Abb80531 Hepatitis
13	45	80.4	11	5 ABB80525	Abb80525 Hepatitis
14	45	80.4	11	5 ABB80561	Abb80561 Hepatitis
15	45	80.4	11	5 ABB80521	Abb80521 Hepatitis
16	45	80.4	11	5 ABB80522	Abb80522 Hepatitis
17	45	80.4	11	5 ABB80566	Abb80566 Hepatitis
18	45	80.4	11	5 ABB80563	Abb80563 Hepatitis
19	45	80.4	11	5 ABB80565	Abb80565 Hepatitis
20	45	80.4	11	5 ABB80524	Abb80524 Hepatitis
21	45	80.4	11	5 ABB80529	Abb80529 Hepatitis
22	45	80.4	11	5 ABB80567	Abb80567 Hepatitis
23	45	80.4	11	5 ABB80528	Abb80528 Hepatitis
24	45	80.4	11	5 ABB80562	Abb80562 Hepatitis
25	45	80.4	11	5 ABB80559	Abb80559 Hepatitis

26	45	80.4	11	5 ABB80526	Abb80526 Hepatitis
27	45	80.4	11	5 ABB80564	Abb80564 Hepatitis
28	45	80.4	11	5 ABB80568	Abb80568 Hepatitis
29	40	71.4	11	5 ABB80536	Abb80536 Hepatitis
30	40	71.4	11	5 ABB80542	Abb80542 Hepatitis
31	40	71.4	11	5 ABB80543	Abb80543 Hepatitis
32	40	71.4	11	5 ABB80535	Abb80535 Hepatitis
33	40	71.4	11	5 ABB80538	Abb80538 Hepatitis
34	40	71.4	11	5 ABB80540	Abb80540 Hepatitis
35	40	71.4	11	5 ABB80539	Abb80539 Hepatitis
36	39	69.6	11	5 ABB80548	Abb80548 Hepatitis
37	39	69.6	11	5 ABB80549	Abb80549 Hepatitis
38	39	69.6	11	5 ABB80547	Abb80547 Hepatitis
39	39	69.6	11	5 ABB80544	Abb80544 Hepatitis
40	39	69.6	11	5 ABB80556	Abb80556 Hepatitis
41	39	69.6	11	5 ABB80557	Abb80557 Hepatitis
42	39	69.6	11	5 ABB80551	Abb80551 Hepatitis
43	39	69.6	11	5 ABB80553	Abb80553 Hepatitis
44	39	69.6	11	5 ABB80552	Abb80552 Hepatitis
45	39	69.6	11	5 ABB80545	Abb80545 Hepatitis

ALIGNMENTS

RESULT 1

ABB80523	ID	ABB80523	standard; peptide; 11 AA.
XX	AC	ABB80523;	
XX	DT	08-OCT-2002	(first entry)
XX	DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.	
XX	KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.	
XX	OS	Synthetic.	
XX	FH	Key	Location/Qualifiers
FT	Modified-site	1	/note= "N-terminal acetyl"
FT	Modified-site	6	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference	9	/note= "D-form residue"
FT	Modified-site	11	/note= "C-terminal amide"
FT	WO200208251-A2.		
PD	31-JAN-2002.		
XX	19-JUL-2001;	2001WO-US023169.	
XX	21-JUL-2000;	2000US-0220101P.	
XX	(CORV-)	CORVAS INT INC.	
XX	Lim-Wilby M,	Levy OE, Brunck TK;	
XX	WPI;	2002-361643/39.	
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.		
PS	Claim 17;	Page 64; 69pp; English.	
XX	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the		

CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMHYS 11

RESULT 2
ABB80558
ID ABB80558 standard; peptide; 11 AA.
XX
AC ABB80558;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KW virucide.
XX
OS Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT Modified-site 11 /note= "C-terminal amide"
XX

WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071; Mismatches 0; Indels 0;
Matches 11; Conservative 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMHYS 11

RESULT 3
ABB80560
ID ABB80560 standard; peptide; 11 AA.
XX
AC ABB80560;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KW virucide.
XX
OS Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT Modified-site 11 /note= "C-terminal amide"
XX

WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0;

2Y 1 EEVVPXGMHYS 11
 |||||
 Db 1 EEVVPXGMHYS 11

RESULT 4

ABB80527

ID ABB80527 standard; peptide; 11 AA.

XX AC ABB80527;
 XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.

XX CS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Modified-site 1 /note= "N-terminal acetyl"

XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

XX FT Misc-difference 8

XX FT Modified-site 11 /note= "D-form residue"

XX FT Modified-site 11 /note= "C-terminal amide"

XX PN WO200208251-A2.

XX XX 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX XX (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX XX WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.00071;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11

Db 1 EEVVPXGMHYS 11

RESULT 5

ABB80537

ID ABB80537 standard; peptide; 11 AA.

XX AC ABB80537;
 XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.

XX CS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Modified-site 1 /note= "N-terminal acetyl"

XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

XX FT Misc-difference 9 /note= "D-form residue"

XX FT Modified-site 11 /note= "C-terminal amide"

XX PN WO200208251-A2.

XX XX 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX XX (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX XX WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 87.5%; Score 49; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0069;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11

Db 1 EEVVPXGMHYS 11

RESULT 6

ABB80541

ID ABB80541 standard; peptide; 11 AA.

XX AC ABB80541;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
XX Query Match 85.7%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.011; Indels 0; Gaps 0;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB ||||| |||
1 EEVVPXGSHYS 11
RESULT 9
AB80550
ID AB80550 standard; peptide; 11 AA.
XX AC AB80550;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX PD 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
XX Query Match 85.7%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.011; Indels 0; Gaps 0;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB ||||| |||
1 EEVVPXGSHYS 11
RESULT 10
AB80555
ID AB80555 standard; peptide; 11 AA.
XX AC AB80555;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX PD 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.

XX	Claim 17; Page 65; 69pp; English.	
PS		
XX	The sequence represents a peptide compound of the invention having	
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the	
CC	invention are alpha-ketoamide peptide analogues. The peptides have	
CC	virucide activity, and are useful for treating and in the manufacture of	
CC	a medicament to treat disorders associated with HCV protease. A	
CC	pharmaceutical composition comprising the peptide as an active ingredient	
CC	is useful for treating disorders associated with hepatitis C virus	
XX		
SQ	Sequence 11 AA;	
	Query Match 85.7%; Score 48; DB 5; Length 11;	
	Best Local Similarity 90.9%; Pred. No. 0.011; 1; Indels 0; Gaps 0;	
	Matches 10; Conservative 0; Mismatches 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 EEVVPXGMHYS 11	
DB	1 EEVVPXGSHYS 11	
RESULT 11		
ABB80532		
ID	ABB80532 standard; peptide; 11 AA.	
XX		
AC	ABB80532;	
XX		
DT	08-OCT-2002 (first entry)	
XX		
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.	
XX		
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;	
KW	virucide.	
XX		
OS	Synthetic.	
XX		
FH	Key Location/Qualifiers	
FT	Modified-site 1 /note= "N-terminal acetyl"	
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with	
FT	residue 7"	
FT	Misc-difference 9 /note= "D-form residue"	
FT	Modified-site 11 /note= "C-terminal amide"	
FT		
XX	WO200208251-A2.	
XX		
PD	31-JAN-2002.	
XX		
PF	19-JUL-2001; 2001WO-US023169.	
XX		
PR	21-JUL-2000; 2000US-0220101P.	
XX		
PA	(CORV-) CORVAS INT INC.	
XX		
PI	Lim-Wilby M, Levy OE, Brunck TK;	
XX		
DR	WPI; 2002-361643/39.	
XX		
PF	19-JUL-2001; 2001WO-US023169.	
XX		
PR	21-JUL-2000; 2000US-0220101P.	
XX		
PA	(CORV-) CORVAS INT INC.	
XX		
PI	Lim-Wilby M, Levy OE, Brunck TK;	
XX		
DR	WPI; 2002-361643/39.	
XX		
PF	Novel peptide compound having hepatitis C virus protease inhibitory	
PT	activity useful for treating disorders associated with hepatitis C virus	
PT	protease.	
XX		
PS	Claim 17; Page 64; 69pp; English.	
XX		
CC	The sequence represents a peptide compound of the invention having	
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the	
CC	invention are alpha-ketoamide peptide analogues. The peptides have	
CC	virucide activity, and are useful for treating and in the manufacture of	
CC	a medicament to treat disorders associated with HCV protease. A	
CC	pharmaceutical composition comprising the peptide as an active ingredient	
CC	is useful for treating disorders associated with hepatitis C virus	
XX		
SQ	Sequence 11 AA;	
	Query Match 85.7%; Score 48; DB 5; Length 11;	
	Best Local Similarity 90.9%; Pred. No. 0.011; 1; Indels 0; Gaps 0;	
	Matches 10; Conservative 0; Mismatches 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 EEVVPXGMHYS 11	
DB	1 EEVVPXGSHYS 11	
RESULT 11		
ABB80532		
ID	ABB80532 standard; peptide; 11 AA.	
XX		
AC	ABB80532;	
XX		
DT	08-OCT-2002 (first entry)	
XX		
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.	
XX		
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;	
KW	virucide.	
XX		
OS	Synthetic.	
XX		
FH	Key Location/Qualifiers	
FT	Modified-site 1 /note= "N-terminal acetyl"	
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with	
FT	residue 7"	
FT	Misc-difference 9 /note= "D-form residue"	
FT	Modified-site 11 /note= "C-terminal amide"	
FT		
XX	WO200208251-A2.	
XX		
PD	31-JAN-2002.	
XX		
PF	19-JUL-2001; 2001WO-US023169.	
XX		
PR	21-JUL-2000; 2000US-0220101P.	
XX		
PA	(CORV-) CORVAS INT INC.	
XX		
PI	Lim-Wilby M, Levy OE, Brunck TK;	
XX		
DR	WPI; 2002-361643/39.	
XX		
PF	Novel peptide compound having hepatitis C virus protease inhibitory	
PT	activity useful for treating disorders associated with hepatitis C virus	
PT	protease.	
XX		
PS	Claim 17; Page 64; 69pp; English.	
XX		
CC	The sequence represents a peptide compound of the invention having	
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the	
CC	invention are alpha-ketoamide peptide analogues. The peptides have	
CC	virucide activity, and are useful for treating and in the manufacture of	
CC	a medicament to treat disorders associated with HCV protease. A	
CC	pharmaceutical composition comprising the peptide as an active ingredient	
CC	is useful for treating disorders associated with hepatitis C virus	
XX		
SQ	Sequence 11 AA;	
	Query Match 82.1%; Score 46; DB 5; Length 11;	
	Best Local Similarity 90.9%; Pred. No. 0.027; 1; Indels 0; G	

```

jb      ||||| |||
      1 EEVVPXGMHYS 11

RESULT 13
BB80525
D ABB80525 standard; peptide; 11 AA.
X
X
X ABB80525;
X
X 08-OCT-2002 (first entry)
X
X Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
X
X Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
X virucide.
X
X Synthetic.
X
X Key Location/Qualifiers
X Modified-site 1 /note= "N-terminal acetyl"
X Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
X residue 7"
X Modified-site 8 /note= "Oxymethionine"
X Misc-difference 8 /note= "D-form residue"
X Modified-site 11 /note= "C-terminal amide"
X
X WO200208251-A2.
X
X 31-JAN-2002.
X
X 19-JUL-2001; 2001WO-US023169.
X
X 21-JUL-2000; 2000US-0220101P.
X (CORV-) CORVAS INT INC.
X
X Lim-Wilby M, Levy OE, Brunck TK;
X
X WPI; 2002-361643/39.
X
X Novel peptide compound having hepatitis C virus protease inhibitor
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X
X Claim 17; Page 64; 69pp; English.
X
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus
X
X Query Match 80.4%; Score 45; DB 5; Length 11;
X Best Local Similarity 90.9%; Pred. No. 0.042; 1; Indels 0; Gaps 0;
X Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
X
X 1 EEVVPXGMHYS 11
X ||||| |||
X 1 EEVVPXGMHYS 11
X
X RESULT 14
X BB80561
X D ABB80561 standard; peptide; 11 AA.
X

```

```

AC ABB80561;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Modified-site 8 /note= "Oxymethionine"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
XX Query Match 80.4%; Score 45; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.042; 1; Indels 0; Gaps 0;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 EEVVPXGMHYS 11
XX ||||| |||
XX 1 EEVVPXGMHYS 11
XX
XX RESULT 15
XX ABB80521
XX ID ABB80521 standard; peptide; 11 AA.
XX
XX ABB80521;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX

```

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
FT FH Modified-site 1
FT FT /note= "N-terminal acetyl"
FT FT Modified-site 6
FT FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT FT Modified-site 11
FT FT /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus

SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 5; Length 11;
Best Local Similarity 90.3%; Pred. No. 0.042;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

Search completed: June 3, 2004, 11:48:22
Job time : 45.9333 secs

Result No.	Query #	Score	Match	Length	DB	ID	Description
1	39	69.6	1037	4	US-09-134-001C-4794		Sequence 4794, Ap
2	37	66.1	856	4	US-09-252-991A-21444		Sequence 21444, A
3	34	60.7	323	4	US-09-543-681A-7304		Sequence 7304, Ap
4	34	60.7	600	2	US-08-821-119-19		Sequence 19, Appl
5	34	60.7	600	2	US-08-821-118-2		Sequence 2, Appl
6	33	58.9	277	4	US-09-252-991A-26615		Sequence 26615, A
7	33	58.9	385	4	US-09-252-991A-27834		Sequence 27834, A
8	33	58.9	747	4	US-09-724-864-36		Sequence 36, Appl
9	33	58.9	3472	4	US-09-408-020-4		Sequence 4, Appl
10	32	57.1	70	4	US-09-134-001C-3950		Sequence 3950, Ap
11	32	57.1	101	4	US-09-621-976-6096		Sequence 6096, Ap
12	32	57.1	102	2	US-08-580-988A-23		Sequence 23, Appl
13	32	57.1	126	2	US-08-878-995A-3		Sequence 3, Appl
14	32	57.1	126	3	US-09-215-096-3		Sequence 3, Appl
15	32	57.1	152	2	US-08-460-694-4		Sequence 4, Appl
16	32	57.1	152	3	US-08-460-744-4		Sequence 4, Appl
17	32	57.1	152	3	US-07-667-711B-4		Sequence 4, Appl
18	32	57.1	173	1	US-08-193-977-7		Sequence 7, Appl
19	32	57.1	189	2	US-08-464-517-21		Sequence 21, Appl
20	32	57.1	189	2	US-08-246-361A-21		Sequence 21, Appl
21	32	57.1	189	3	US-08-463-772-21		Sequence 21, Appl
22	32	57.1	189	5	PCT-US93-05000-21		Sequence 21, Appl
23	32	57.1	236	2	US-08-464-517-22		Sequence 22, Appl
24	32	57.1	236	2	US-08-246-361A-22		Sequence 22, Appl
25	32	57.1	236	3	US-08-463-772-22		Sequence 22, Appl
26	32	57.1	236	5	PCT-US93-05000-22		Sequence 22, Appl
27	32	57.1	280	2	US-08-464-517-6		Sequence 6, Appl

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 66.1%; Score 37; DB 4; Length 856;
Best Local Similarity 70.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGMXY 10
Db 64 EAVVPGGXY 73

RESULT 3

US-09-543-681A-7304
Sequence 7304, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 7304
LENGTH: 223
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-7304

Query Match 60.7%; Score 34; DB 4; Length 323;
Best Local Similarity 55.6%; Pred. No. 63;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGMXY 10
Db 75 DVCVPGVHY 83

RESULT 4

US-08-821-119-19
Sequence 19, Application US/08821119
Patent No. 5821104
GENERAL INFORMATION:
APPLICANT: Holm, Kaj Andre
APPLICANT: Rasmussen, Grethe
APPLICANT: Haikier, Torben
APPLICANT: Lehbeck, Jan
TITLE OF INVENTION: Tripeptidyl Aminopeptidase
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5821104 No. 5821104disk of No. 5821104th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,119
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107-204-US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-119-19

Query Match 60.7%; Score 34; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGMHYS 11
Db 31 VPKGMHYS 38

RESULT 5

US-08-821-118-2
Sequence 2, Application US/08821118
Patent No. 5989889
GENERAL INFORMATION:
APPLICANT: Rev. Michael
APPLICANT: Golightly, Elizabeth
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5989889 No. 5989889disk of No. 5989889th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,118
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107.400-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-118-2

Query Match 60.7%; Score 34; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGMHYS 11
Db 31 VPKGMHYS 38

b) 31 VPKGWHYS 38

RESULT 6
US-09-252-991A-26615
Sequence 26615, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26615
LENGTH: 277
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match 58.9%; Score 33; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Y 1 EVVVPXGMHYS 11
|||
48 EETVPGGGHTS 58

RESULT 7
US-09-252-991A-27834
Sequence 27834, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27834
LENGTH: 385
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27834

Query Match 58.9%; Score 33; DB 4; Length 385;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Y 2 EVVVPXGMHY 10
|||
201 EILPALHY 209

RESULT 8
US-09-724-864-36
Sequence 36, Application US/09724864
Patent No. 6380362
GENERAL INFORMATION:
APPLICANT: Watson, James G.
TITLE OF INVENTION: Polynucleotides, polypeptides expressed

; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 11000.1050U1
; CURRENT APPLICATION NUMBER: US/09/724,864
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Rat
US-09-724-864-36

Query Match 58.9%; Score 33; DB 4; Length 747;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMHYS 11
|||
Db 627 PGLRHS 633

RESULT 9
US-09-408-020-4
Sequence 4, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCRP-002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3472
TYPE: PRT
ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 58.9%; Score 33; DB 4; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.4e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMHYS 11
|||
Db 2294 EDVIPRGISFS 2304

RESULT 10
US-09-134-001C-3950
Sequence 3950, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3950
LENGTH: 70
TYPE: PRT

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3950

Query Match 57.1%; Score 32; DB 4; Length 70;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGMHYS 11
DB 36 MPKGFHYS 43

RESULT 11

US-09-621-976-6096
Sequence 6096, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent.pm

SEQ ID NO 6096

LENGTH: 101

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: UNSURE

LOCATION: 96

OTHER INFORMATION: Xaa = *, Ala, Glu, Gly, Ile, Lys, Leu, Arg, Ser, Thr, Val
US-09-621-976-6096

Query Match 57.1%; Score 32; DB 4; Length 101;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMHY 10
DB 40 PRGMHY 45

RESULT 12

US-08-580-988A-23
Sequence 23, Application US/08580988A
Patent No. 5856161
GENERAL INFORMATION:

APPLICANT: Aggarwal et al.
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-1-Associated Protein Kinase And Methods
TITLE OF INVENTION: For Its Use
NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dr. Benjamin A. Adler

STREET: 8011 Candle Lane

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77071

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 Mb floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Microsoft Word for Macintosh

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/580,988A

FILING DATE: January 3, 1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5721CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 102 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: protein

HYPOTHETICAL: no

ANTI-SENSE: no

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

US-08-580-988A-23

Query Match 57.1%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
DB 24 BEVFPPLAMNY 33

RESULT 13

US-08-879-995A-3
Sequence 3, Application US/08879995A
Patent No. 5985606
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Kaser, Matthew R.

TITLE OF INVENTION: HUMAN PREPROTACHYKININ B

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/879,995A

FILING DATE: Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0326 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
JS-08-879-995A-3

Query Match 57.1%; Score 32; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
|:|:|:|:|
Db 28 EQVPGGCH 36

RESULT 14

JS-09-215-096-3
Sequence 3, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
JS-09-215-096-3

Query Match 57.1%; Score 32; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
|:|:|:|:|
Db 28 EQVPGGCH 36

RESULT 15

US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5858655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,694
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609.4070002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-694-4

Query Match 57.1%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 10
|:|:|:|:|
Db 20 BEVFPPLAMNY 29

Search completed: June 3, 2004, 12:03:06
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)

91.741 Million cell updates/sec

Title: US-09-909-164-7

Perfect score: 56

Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	54	96.4	11	12	US-09-909-164-7	Sequence 7, Appl
2	54	96.4	11	12	US-09-909-164-11	Sequence 11, Appl
3	49	87.5	11	12	US-09-909-164-21	Sequence 21, Appl
4	49	87.5	11	12	US-09-909-164-25	Sequence 25, Appl
5	48	85.7	11	12	US-09-909-164-30	Sequence 30, Appl
6	48	85.7	11	12	US-09-909-164-34	Sequence 34, Appl
7	48	85.7	11	12	US-09-909-164-38	Sequence 38, Appl
8	48	85.7	11	12	US-09-909-164-39	Sequence 39, Appl
9	48	85.7	11	12	US-09-909-164-42	Sequence 42, Appl
10	48	85.7	11	12	US-09-909-164-44	Sequence 44, Appl
11	46	82.1	11	12	US-09-909-164-15	Sequence 15, Appl
12	46	82.1	11	12	US-09-909-164-16	Sequence 16, Appl
13	45	80.4	11	12	US-09-909-164-5	Sequence 5, Appl
14	45	80.4	11	12	US-09-909-164-6	Sequence 6, Appl
15	45	80.4	11	12	US-09-909-164-8	Sequence 8, Appl

Sequence 9, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 47, Appl
Sequence 48, Appl
Sequence 49, Appl
Sequence 50, Appl
Sequence 51, Appl
Sequence 52, Appl
Sequence 19, Appl
Sequence 20, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 24, Appl
Sequence 26, Appl
Sequence 27, Appl
Sequence 28, Appl
Sequence 29, Appl
Sequence 31, Appl
Sequence 32, Appl
Sequence 33, Appl
Sequence 35, Appl
Sequence 36, Appl
Sequence 37, Appl
Sequence 40, Appl
Sequence 41, Appl
Sequence 43, Appl
Sequence 45, Appl
Sequence 46, Appl

US-09-909-164-9
US-09-909-164-10
US-09-909-164-12
US-09-909-164-13
US-09-909-164-47
US-09-909-164-48
US-09-909-164-49
US-09-909-164-50
US-09-909-164-51
US-09-909-164-52
US-09-909-164-19
US-09-909-164-20
US-09-909-164-22
US-09-909-164-23
US-09-909-164-24
US-09-909-164-26
US-09-909-164-27
US-09-909-164-28
US-09-909-164-31
US-09-909-164-32
US-09-909-164-33
US-09-909-164-35
US-09-909-164-36
US-09-909-164-37
US-09-909-164-40
US-09-909-164-41
US-09-909-164-43
US-09-909-164-45
US-09-909-164-46

ALIGNMENTS

RESULT 1

US-09-909-164-7

; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)

```
; OTHER INFORMATION: AMIDATION
US-09-909-164-7
Query Match      96.4%; Score 54; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 2
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-21
Query Match      87.5%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0067;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 4
US-09-909-164-25
; Sequence 25, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 25
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

Query Match      96.4%; Score 54; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 3
US-09-909-164-21
; Sequence 21, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
```

OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
S-09-909-164-25

Query Match 87.5%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0067;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 EEVVPXGMHYS 11
| | | | | | | | | |
b 1 EEVVPXGMHYS 11

RESULT 5

US-09-909-164-30

Sequence 30, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Bruck, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

CURRENT FILING DATE: 2003-03-25

PRIOR APPLICATION NUMBER: 60/220,101

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 30

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:

NAME/KEY: MOD RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)

IS-09-909-164-30

Query Match 85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 EEVVPXGMHYS 11
| | | | | | | | | |
b 1 EEVVPXGMHYS 11

RESULT 6

US-09-909-164-34

Sequence 34, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Bruck, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 34
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLTATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
OTHER INFORMATION: 11-mer synthesized according to example 1
US-09-909-164-34

Query Match 85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMHYS 11

RESULT 7

US-09-909-164-38

Sequence 38, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Bruck, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

CURRENT FILING DATE: 2003-03-25

PRIOR APPLICATION NUMBER: 60/220,101

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 38

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:

NAME/KEY: MOD RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (8)..(8)

```
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
JS-09-909-164-38

Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y 1 BEVVPXGMHYS 11
   |||||
Db 1 BEVVPXGSHYS 11

RESULT 8
US-09-909-164-39
; Sequence 39, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
US-09-909-164-42

Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 BEVVPXGMHYS 11
   |||||
Db 1 BEVVPXGSHYS 11

RESULT 10
US-09-909-164-44
; Sequence 44, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acids
US-09-909-164-39

Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 BEVVPXGMHYS 11
   |||||
Db 1 BEVVPXGSHYS 11

RESULT 9
US-09-909-164-42
; Sequence 42, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
```

RESULT 12
IS-09-909-164-16
Sequence 16. Application US/09909164

; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-5

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
||| ||||| ||
Db 1 BEVVPXGMHYS 11

RESULT 14

US-09-909-164-6
; Sequence 6, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-6

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
||| ||||| ||
Db 1 BEVVPXGMHYS 11

RESULT 15

US-09-909-164-8
; Sequence 8, Application US/09909164

; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-8

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
||| ||||| ||
Db 1 BEVVPXGMHYS 11

Search completed: June 3, 2004, 12:57:15
Job time : 33.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-7

Perfect score: 56

Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	66.1	308	2 A72207	ftsH proteinase ac
2	37	66.1	1057	2 F98992	carbamoyl-phosphat
3	36	64.3	102	2 A42452	V1 protein - tobac
4	36	64.3	252	2 A22001	hypothetical prote
5	36	64.3	460	2 S89046	hypothetical prote
6	36	64.3	743	2 S38143	hypothetical prote
7	35	62.5	156	2 D82618	conserved hypothet
8	35	62.5	233	2 T02590	DNA binding protei
9	35	62.5	311	2 H69194	GMP synthetase, su
10	35	62.5	425	2 T24111	hypothetical prote
11	35	62.5	510	2 G86430	T518.1 protein - A
12	34	60.7	264	2 G93117	diphthine synthase
13	34	60.7	279	2 C75538	hypothetical prote
14	34	60.7	350	2 B75478	3-dehydroquinatase
15	34	60.7	355	2 T35025	probable DNA ligas
16	34	60.7	360	2 E69086	cell division prot
17	34	60.7	425	2 C83903	hypothetical prote
18	34	60.7	426	2 S58132	Slsl protein precu
19	34	60.7	495	2 T28717	hypothetical prote
20	34	60.7	1028	2 A73286	ATP-dependent DNA
21	33	58.9	156	2 S54619	hypothetical prote
22	33	58.9	367	2 E83607	polyamine transpor
23	33	58.9	441	2 G82253	conserved hypothet
24	33	58.9	466	2 G71542	probable amino aci
25	33	58.9	466	2 H81697	amino acid antipor
26	33	58.9	487	2 S58811	finger protein (Cl
27	33	58.9	514	1 HQDVLB	cytochrome-c3 hydr
28	33	58.9	534	2 A69284	coenzyme F420-quin
29	33	58.9	545	2 T08564	hypothetical prote

DNA mismatch repai
macrophage-stimula
C14B9.8 protein -
L-shaped tail fibe
hypothetical 367K
tachykinin B precu
heme exporter prot
hypothetical prote
ribosomal protein
probable ThUA prot
hypothetical prote
cyclin D2 - rat
cyclin D2 - mouse
cyclin D2 - human

ALIGNMENTS

RESULT 1

A72207

ftsH proteinase activity modulator HflK - Thermotoga maritima (strain MSB8)

C.Species: Thermotoga maritima

C.Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C.Accession: A72207

R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.

Nature 399, 323-329, 1999

A.Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A.Reference number: A72200; MUID:99287316; PMID:10360571

A.Accession: A72207

A>Status: Preliminary

A.Molecule type: DNA

A.Residues: 1-308 <ARN>

A.Cross-references: GB:AE001819; GB:AE000512; NID:g4982396; PIDN:AAD36885.1; PID:g49824

A.Experimental source: strain MSB8

C.Genetics:

A.Gene: TM1822

C.Superfamily: erythrocyte band 7 integral membrane protein

Query Match 66.1% Score 37; DB 2; Length 308;

Best Local Similarity 75.0% Pred. No. 10;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10

DB -41 VVP8GIHY 48

RESULT 2

F98992

carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N31)

C.Species: Staphylococcus aureus

C.Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C.Accession: F98992

R.Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A.Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A.Reference number: A89758; MUID:21311952; PMID:11418146

A.Accession: F98992

A>Status: Preliminary

A.Molecule type: DNA

A.Residues: 1-1057 <KUR>

A.Cross-references: GB:BA000018; PID:g13701002; PIDN:BA842298.1; GSPDB:GN00149

A.Experimental source: strain N315

C.Genetics:

A.Gene: PyrAB

C.Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 66.1%; Score 37; DB 2; Length 1057;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMHYS 11
:|:|:|:|:|:|
Db 190 EIVSNGLHY 199

RESULT 3
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virolgy 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92198538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 64.3%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMHYS 11
:|:|:|:|:|:|
Db 7 QVPSGINYS 16

RESULT 4
A82001
Hypothetical protein alr1563 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: A82001
R:Kaneok, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: A81807; MUID:21595285; PMID:11759840
A:Accession: A82001
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA877929.1; PID:g17135383; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1563

Query Match 64.3%; Score 36; DB 2; Length 252;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEWVPXGMHY 10
:|:|:|:|:|:|
Db 235 EMIVPAGLHF 244

RESULT 5
S69046
Hypothetical protein YPL139c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 17-Mar-2000
C:Accession: S69046
R:Hall, J.; Depaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040
A:Accession: S69046
A:Molecule type: DNA
A:Residues: 1-460 <HAL>
A:Cross-references: EMBL:U43703; NID:g1244769; PIDN:AAB68221.1; PID:g1244776; MIPS:YPL1
C:Genetics:
A:Gene: SGD:UWE1
A:Cross-references: SGD:S0006060; MIPS:YPL139c
A:Map position: 16L
C:Superfamily: Saccharomyces cerevisiae transcription modulator WTM1

Query Match 64.3%; Score 36; DB 2; Length 460;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
:|:|:|:|:|:|
Db 85 IVPLGLHY 92

RESULT 6
S38143
Hypothetical protein YBL011w homolog YKR067w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-Apr-2002
C:Accession: S38143
R:van Vliet-Reedijk, J.C.; Planta, R.J.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38130
A:Accession: S38143
A:Molecule type: DNA
A:Residues: 1-743 <VAN>
A:Cross-references: EMBL:Z28292; NID:g486536; PIDN:CMA82146.1; PID:g486537; MIPS:YKR067
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GPT2
A:Cross-references: SGD:S0001775
A:Map position: 11R
C:Keywords: transmembrane protein

Query Match 64.3%; Score 36; DB 2; Length 743;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
:|:|:|:|:|:|
Db 294 VVPCGLHY 301

RESULT 7
D82618
conserved hypothetical protein XPI1950 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82618
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82618
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <SIM>
A:Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GN00
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canaro, L.E.A.; Carraro, D.M.; Carrer,
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fro
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laic
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,

A;Authors: Martins, E.M.F.; Matsuikuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Falmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1950

Query Match 62.5%; Score 35; DB 2; Length 156;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
|||:|:|
Db 119 BEILPQGVH 127

RESULT 8
T02590
DNA binding protein EREP-2 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
C;Accession: T02590
R;Ohme-Takagi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi
A;Reference number: Z14671; MUID:95276459; PMID:7756828
A;Accession: T02590
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-233 <OHM>
A;Cross-references: EMBL:D38126; NID:G790362; PIDN:BA07324.1; PID:gl208498
A;Experimental source: strain B4; tissue-type leaf

Query Match 62.5%; Score 35; DB 2; Length 233;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 10
:|:|:|
Db 90 QAVVPKGRHY 99

RESULT 9
H69194
GMP synthetase, subunit B - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: H69194
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: H69194
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-311 <MTH>
A;Cross-references: GB:AE000850; GB:AE000666; NID:g2621794; PIDN:AA85215.1; PID:g262179
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH710
A;Start codon: GTG

Query Match 62.5%; Score 35; DB 2; Length 311;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 11
|||:|:|

Db 219 BEVVESGLHES 229

RESULT 10
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111
R;Percy, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19842
A;Accession: T24111
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CBSP:R10D12.10
A;Experimental source: clone R10D12
C;Genetics:
A;Gene: CBSP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 62.5%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 10
|:|:|:|
Db 335 EQIVPGLQY 344

RESULT 11
G86430
TS18.1 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C;Accession: G86430
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86430
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-510 <STO>
A;Cross-references: GB:AE005172; NID:g4587512; PIDN:AA25743.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: hexose phosphate transport protein uhpt

Query Match 62.5%; Score 35; DB 2; Length 510;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 10
|||:|:|
Db 12 BEVVPGLHF 21

RESULT 12
G69117
diphthine synthase - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 25-Aug-2003
C;Accession: G69117

R.; Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicarte, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
 A;Reference number: A59000; MUID:98037514; PMID:9371463
 A;Accession: G69117
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-264 <MTH>
 A;Cross-references: GB:AE000940; GB:AE000666; NID:G2623011; PIDN:AB86340.1; PID:G262301
 A;Experimental source: strain Delta H
 C;Genetics:
 A;Gene: MTH1874
 C;Superfamily: diphthamide biosynthesis methyltransferase

Query Match 60.7%; Score 34; DB 2; Length 264;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMHY 10
 |||||
 Db 235 VVPAGLHF 242
 |||||

RESULT 13
 C75538
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C;Accession: C75538
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: C75538
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-279 <WHI>
 A;Cross-references: GB:AE001889; GB:AE000513; NID:G6457944; PIDN:AAF09867.1; PID:G645795
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0271
 A;Map position: 1
 C;Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 60.7%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 VPXGMHY 11
 |||||
 Db 100 VPLGRHYS 107
 |||||

RESULT 14
 B75478
 3-dehydroquinate synthase - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C;Accession: B75478
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: B75478
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-350 <WHI>

A;Cross-references: GB:AE001932; GB:AE000513; NID:G6458481; PIDN:AAF10353.1; PID:G64584
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0777
 C;Superfamily: 3-dehydroquinate synthase; 3-dehydroquinate synthase homology

Query Match 60.7%; Score 34; DB 2; Length 350;
 Best Local Similarity 60.0%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPXGMHY 11
 |||||
 Db 252 EAVAYGMHYA 261
 |||||

RESULT 15
 T35025
 Probable DNA ligase - Streptomyces coelicolor
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C;Accession: T35025
 R;Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A;Reference number: Z21565
 A;Accession: T35025
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-355 <SEE>
 A;Cross-references: EMBL:AL079355; PIDN:CAB45581.1; GSPDB:GN00070; SCOEDB:SC4C6.17C
 A;Experimental source: strain A3(2)
 C;Genetics:
 A;Gene: SCOEDB:SC4C6.17C

Query Match 60.7%; Score 34; DB 2; Length 355;
 Best Local Similarity 71.4%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 VPXGMHY 10
 |||||
 Db 20 IPPGMHY 26
 |||||

Search completed: June 3, 2004, 11:59:59
 Job time : 10 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds

(without alignments)

117.693 Million cell updates/sec

Title: US-09-909-164-7

Perfect score: 56

Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	69.6	1057	1	Q8CPJ4 staphylococ
2	38	67.9	1058	1	Q8986 fuscobacteri
3	37	66.1	426	1	Q87XQ9 vibrio para
4	37	66.1	1057	1	Q9UR5 staphylococ
5	37	66.1	1057	1	P58940 staphylococ
6	36	64.3	102	1	P31619 tobaccos yel
7	36	64.3	460	1	Q03010 saccharomyc
8	36	64.3	743	1	P36148 saccharomyc
9	35	62.5	227	1	O35586 mesocricetu
10	35	62.5	308	1	O26806 methanobact
11	34	60.7	264	1	Q27902 methanobact
12	34	60.7	426	1	Q99158 yarowia li
13	33	58.9	441	1	P46231 vibrio para
14	33	58.9	513	1	P13065 desulfovibr
15	33	58.9	627	1	P49850 bacillus su
16	33	58.9	1188	1	P34335 caenorhabdi
17	33	58.9	1396	1	P13390 bacterioph
18	32	57.1	126	1	P08858 bos taurus
19	32	57.1	153	1	O76217 anopheles g
20	32	57.1	212	1	Q878W6 vibrio para
21	32	57.1	233	1	Q97166 clostridium
22	32	57.1	267	1	P56351 chlorella v
23	32	57.1	288	1	Q04827 rattus norv
24	32	57.1	289	1	P30279 homo sapien
25	32	57.1	289	1	P30280 mus musculu
26	32	57.1	291	1	Q90459 brachydanio
27	32	57.1	291	1	P50755 xenopus lae
28	32	57.1	291	1	P49706 gallus gall
29	32	57.1	291	1	P53782 xenopus lae
30	32	57.1	292	1	P55169 gallus gall
31	32	57.1	292	1	CGD1_CHICK
32	32	57.1	295	1	CGD3_HUMAN
33	32	57.1	295	1	CGD1_HUMAN
34	32	57.1	295	1	CGD1_MOUSE

34	32	57.1	295	1	CGD1_RAT	P39948 rattus norv
35	32	57.1	341	1	HYPE_AZOVI	P40595 azotobacter
36	32	57.1	353	1	T2BA_BACAR	P19887 bacillus an
37	32	57.1	573	1	SUOX_DROME	Q9VWP4 drosophila
38	32	57.1	578	1	MDLB_BUCBP	Q89A96 buchnera ap
39	32	57.1	759	1	SC11_YEAST	P32784 saccharomyc
40	32	57.1	877	1	SULH_SCHPO	O74377 schizosacch
41	32	57.1	1401	1	RPOC_VIBCH	Q9KV29 vibrio chol
42	32	57.1	2717	1	ZEPI_HUMAN	P15822 homo sapien
43	31.5	56.2	847	1	CD22_HUMAN	P20273 homo sapien
44	31	55.4	124	1	REV_SIVCZ	P17280 chimpanzee
45	31	55.4	130	1	SZ05_RAT	P97885 rattus norv

ALIGNMENTS

RESULT 1

ID	CARB	STAEF	STANDARD	PRT	1057 AA
AC	Q8CPJ4				
DT	15-MAR-2004	(Rel. 43, Created)			
DT	15-MAR-2004	(Rel. 43, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).				
GN	CARB OR S50879.				
OS	Staphylococcus epidermidis.				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI_TaxID=1282;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 12228;				
RX	PubMed=12950922;				
RA	Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,				
RA	Qin Z.-Q., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,				
RA	Yuan Z.-H., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,				
RT	"Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228).";				
RL	Mol. Microbiol. 49:1577-1593 (2003).				
CC	-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.				
CC	-!- COPACTOR: Binds 3 manganese ions per subunit (By similarity).				
CC	-!- PATHWAY: Arginine biosynthesis.				
CC	-!- PATHWAY: Pyrimidine biosynthesis; first step.				
CC	-!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).				
CC	-!- SIMILARITY: Belongs to the carb family.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; A016746; AAC00476.1; -.				
DR	HAMAP; MF 01210; -; 1.				
DR	InterPro; IPR006275; CarA_L_glu.				
DR	InterPro; IPR005483; CPase_L.				
DR	InterPro; IPR005479; CPase_L_D2.				
DR	InterPro; IPR005480; CPase_L_D3.				
DR	InterPro; IPR005481; CPase_L_N.				
DR	InterPro; IPR004382; MGS_Like.				
DR	InterPro; IPR000189; SHprot_acsite.				
DR	Pfam; PF00289; CPase_L_chain; 2.				
DR	Pfam; PF02786; CPase_L_D2; 2.				
DR	Pfam; PF02787; CPase_L_D3; 1.				
DR	Pfam; PF02142; MGS; 1.				
DR	PRINTS; PR00098; CPASE.				

DR TIGR01369; CPASaseII_lrg; 1.
 DR PROSITE; PS00866; CPASase 1; 2.
 DR INTERPRO; IPR006275; CarA_L_glu.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 829 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 830 1057 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1057
 FT NP BIND 153 210
 FT NP BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;

 Query Match 69.6%; Score 39; DB 1; Length 1057;
 Best Local Similarity 63.6%; Pred. No. 7.3;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 1 EYVVPXGMHYS 11
 Db 189 KEVSVNGLHYS 199

 RESULT 2
 ID CARB_FUSNN STANDARD; PRT; 1058 AA.
 AC Q8RG66;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 DE CARB OR FN0422.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TAXID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Berman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan F., Chaga O., Goltzman S., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haeelkorn R.,
 RA Fonstein M., Kyripides N., Overbeek R.,
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018 (2002).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
 CC -1- SIMILARITY: Belongs to the carb family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE010554; AA194625.1; ALT_INIT.
 DR HAMAP; MF 01210; 1.
 DR INTERPRO; IPR006275; CarA_L_glu.
 DR INTERPRO; IPR005483; CPase_L_D2.
 DR INTERPRO; IPR005479; CPase_L_D2.
 DR INTERPRO; IPR005480; CPase_L_D3.
 DR INTERPRO; IPR005481; CPase_L_N.
 DR INTERPRO; IPR004362; MGS like.
 DR Pfam; PF02786; CPasease_L_D2; 2.
 DR Pfam; PF02787; CPasease_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPASASE
 DR TIGR01369; CPASaseII_lrg; 1.
 DR PROSITE; PS00866; CPASase 1; 2.
 DR PROSITE; PS00867; CPASase 2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 829 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 830 1057 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1057
 FT NP BIND 153 210
 FT NP BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

 Query Match 67.9%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 QY 2 EYVVPXGMHYS 11
 Db 190 EIVPGLNYS 199

 RESULT 3
 ID AROA_VIBPA STANDARD; PRT; 426 AA.
 AC Q870X9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-enolpyruvylshikimate-3-phosphate synthase) (BPS synthase) (BPSPS).
 DE AROA OR VP1020.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TAXID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae."
 RL Lancet 361:1743-749 (2003).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate = phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -!- SIMILARITY: Belongs to the EPPS synthase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; AP005076; BAC59283.1; -;
 DR HAMAP; MF 00210; -; 1;
 DR InterPro; IPR001986; EPPS_synth.
 DR Pfam; PF0275; EPPS_synthase; 1.
 DR PROSITE; PS00104; EPPS_SYNTHASE_1; 1.
 DR PROSITE; PS00885; EPPS_SYNTHASE_2; 1.
 KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 426 AA; 46094 MW; 373D39C5BA1F70F CRC64;

 Query Match 66.1%; Score 37; DB 1; Length 426;
 Best Local Similarity 60.0%; Pred. No. 7.3;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

 QY 1 EVVPEXGMHY 10
 | : | : | : |
 Db 223 EVVIPAGQHY 232

 RESULT 4
 CARB_STAAM STANDARD; PRT; 1057 AA.
 ID CARB_STAAM STANDARD; PRT; 1057 AA.
 AC Q99UR5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
 DE phosphate synthetase ammonia chain).
 GN CARB OR PYRAB OR SAV1203 OR SA1046.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878, 158879;
 RN [1]
 SEQUENCE FROM N.A.
 RP Lancel 357:1225-1240(2001).
 RC STRAIN=Mu50 / ATCC 700699, and N315;
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Negai Y., Iian J.-Q., Ito T.,
 RA Kanamori M., Matsumori H., Maruyama A., Murakami H., Hoshizawa A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kashiwa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus".
 RL Lancel 357:1225-1240(2001).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the carb family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; AP003361; BAB57365.1; -;
 DR EMBL; AP003132; BAB42298.1; -;
 DR PIR; P89892; P89892.
 DR HSSP; P00968; LCS0.
 DR HAMAP; MF_01210; -; 1;
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR005483; CPhase_L.
 DR InterPro; IPR005479; CPhase_L_D2.
 DR InterPro; IPR005480; CPhase_L_D3.
 DR InterPro; IPR005481; CPhase_L_N.
 DR InterPro; IPR004362; MGS-like.
 DR Pfam; PF00289; CPhase_L_Chain; 2.
 DR Pfam; PF02786; CPhase_L_D2; 2.
 DR Pfam; PF02787; CPhase_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPSASE.
 DR TIGRfams; TIGR01369; CPSaseII_1g; 1.
 DR PROSITE; PS00865; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401
 FT DOMAIN 402 546
 FT DOMAIN 547 929
 FT DOMAIN 930 1057
 FT REPEAT 1 546
 FT REPEAT 547 1057
 FT NP_BIND 153 210
 FT NP_BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

 Query Match 66.1%; Score 37; DB 1; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 2 EVVPEXGMHY 11
 | : | : | : |
 Db 190 EVVNGLHY 199

 RESULT 5
 CARB_STAAM STANDARD; PRT; 1057 AA.
 ID CARB_STAAM STANDARD; PRT; 1057 AA.
 AC P59940;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
 DE phosphate synthetase ammonia chain).
 GN CARB OR PYRAB OR MW1086.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=22040717; PubMed=12044378;
 RX Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RX Nagai Y., Iwana N., Asano K., Naimi T., Kuroda H., Cui L.,
 RX Yamamoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 RT acquired MRSA".
 RL Lancel 359:1819-1827(2002).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.

CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (by
 CC similarity).
 CC
 CC -1- SIMILARITY: Belongs to the carB family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP004825; BAB94951.1; -.
 CC HAMAP; MF 01210; -; 1.
 CC InterPro; IPR006275; CarA_L_glu.
 CC InterPro; IPR005483; CPase_L.
 CC InterPro; IPR005479; CPase_L_D2.
 CC InterPro; IPR005480; CPase_L_D3.
 CC InterPro; IPR005481; CPase_L_N.
 CC InterPro; IPR004362; MGS-like.
 CC Pfam; PF00289; CPase_L_Chain; 2.
 CC Pfam; PF02786; CPase_L_D2; 2.
 CC Pfam; PF02787; CPase_L_D3; 1.
 CC Pfam; PF02142; MGS; 1.
 CC PRINTS; PR00098; CPASE.
 CC TIGRPFAMs; TIGR01369; CPaseII_lrg; 1.
 CC PROSITE; PS00866; CPASE_1; 2.
 CC PROSITE; PS00867; CPASE_2; 2.
 CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 CC ATP-binding; Manganese; Complete proteome.
 CC FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 CC FT DOMAIN 402 545 OLIGOMERIZATION DOMAIN.
 CC FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 CC FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
 CC FT REPEAT 1 546
 CC REPEAT 547 1057
 CC NP_BIND 153 210 ATP (POTENTIAL).
 CC NP_BIND 302 352 ATP (POTENTIAL).
 CC METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 CC METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 CC METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 CC METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 CC SEQUENCE 1057 AA; 117185 MW; DBE3B09F9BC6F152 CRC64;
 CC
 CC Query Match 66.1%; Score 37; DB 1; Length 1057;
 CC Best Local Similarity 60.0%; Pred. No. 19;
 CC Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 2 EVVPGMRYHS 11
 CC :|||:|:
 CC Db 190 EIVSNGLHYS 199
 CC
 CC RESULT 6
 CC Y1LK TYDVA
 CC ID Y1LK TYDVA STANDARD; PRT; 102 AA.
 CC AC P31619;
 CC DT 01-JUL-1993 (Rel. 26, Created)
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)
 CC DT 01-OCT-1993 (Rel. 27, Last annotation update)
 CC DE Hypothetical 11.2 kDa protein.
 CC VI.
 CC OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 CC OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 CC ON NCBI_TaxID=31599;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=92188538; PubMed=1546458;
 CC RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;

RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants.",
 RL Virology 187:633-642(1992).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M81103; AAA47947.1; -.
 CC PIR; A42452; A42452.
 CC InterPro; IPR002621; Gemini_mov.
 CC Pfam; PF01708; Gemini_mov; 1.
 CC Hypothetical protein.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF5B67 CRC64;
 CC
 CC Query Match 64.3%; Score 36; DB 1; Length 102;
 CC Best Local Similarity 60.0%; Pred. No. 2.7;
 CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 EVVPGMRYHS 11
 CC :|||:|:
 CC Db 7 QVVPSSGINS 16
 CC
 CC RESULT 7
 CC UMEL YEAST
 CC ID UMEL YEAST STANDARD; PRT; 460 AA.
 CC AC Q03010; P87330;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Meiosis negative regulator UMEL.
 CC DE UMEL OR WTM3 OR YPL139C OR LPI7C.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=A364A;
 CC Mallory M.J., Strich R.;
 CC Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=S288C / AB972;
 CC MEDLINE=97313271; PubMed=9169875;
 CC Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
 CC Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 CC Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 CC Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 CC Dietrich F.S., Delius H., Dipalo I., Dubois E., Duesterhoeft A.,
 CC Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 CC Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
 CC Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
 CC Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 CC Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
 CC Mueller-Auer S., Namath A., Newtich U., Oefner P., Pearson D.,
 CC Patel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 CC Scherens B., Schramm S., Schroeder M., Sedu A.M., Tettelin H.,
 CC Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 CC Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 CC Zhong W.W., Zollner A., Vo D.H., Hani J.;
 CC "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
 CC Nature 387:103-105(1997)
 CC RT
 CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
 CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
 CC -1- SIMILARITY: Contains 4 WD repeats.
 CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.

FT ACT_SITE 86 86 BY SIMILARITY.
 FT ACT_SITE 148 148 BY SIMILARITY.
 FT SITE 225 227 MICROBODY TARGETING SIGNAL.
 SQ SEQUENCE 227 AA; 26317 MW; F50A6586385B803 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 227;
 Best Local Similarity 70.0%; Pred. No. 9.8;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10

Db 121 BEVDNEMHY 130

RESULT 10

GAAB_METHH STANDARD; PRT; 308 AA.
 AC O26806;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP synthase).
 DE SYNTHASE.
 GN GAAB OR MTH710.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 H₂O = AMP + diphosphate + GMP + L-glutamate.
 CC -!- PATHWAY: GMP biosynthesis.
 CC -!- SUBUNIT: Heterodimer composed of a glutamine amidotransferase
 subunit (A) and a GMP synthase subunit (B) (potential).
 CC -!- SIMILARITY: Belongs to the GMP synthase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AE000850; AAB85215.1; ALT_INIT.

DR HSP; P04079; 1GPM.

DR HAMAP; MF 00345; -; 1.

DR InterPro; IPR001674; GMP synth C.

DR Pfam; PF00958; GMP_synth_C; 1.

DR TIGRfams; TIGR00884; guaA_Cterm; 1.

KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;

Complete proteome.

FT DOMAIN 33 184 GMP-BINDING (BY SIMILARITY).

FT NP_BIND 29 35 ATP (BY SIMILARITY).

FT SEQUENCE 308 AA; 34403 MW; F2DCP6ED202CAEC1 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 308;
 Best Local Similarity 63.6%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 11
 Db 216 BEVVGSLHES 226

RESULT 11

DPHE_METHH STANDARD; PRT; 264 AA.
 AC O27902;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable diphthine synthase (EC 2.1.1.98) (Diphthamide biosynthesis
 methyltransferase).
 DE METHYLTRANSFERASE.
 GN DPHE OR MTH1874.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: Required for the methylation step in diphthamide
 biosynthesis (By similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 2-(3-carboxy-3-
 aminopropyl)-L-histidine = S-adenosyl-L-homocysteine + 2-[3-
 carboxy-3-(methylammonio)propyl]-L-histidine.
 CC -!- PATHWAY: Diphthamide biosynthesis; second step.
 CC -!- SIMILARITY: Belongs to the diphthine synthase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AE000840; AAB86340.1; -.

DR PIR; G69117; G69117.

DR HAMAP; MF 01084; -; 1.

DR InterPro; IPR000878; Cor/por_Mettransf.

DR InterPro; IPR004551; Dphthm_synthase.

DR Pfam; PF00590; TP_methylase; 1.

DR TIGRfams; TIGR00532; dphs; 1.

KW Transferase; Methyltransferase; Complete proteome.

SQ SEQUENCE 264 AA; 28858 MW; 368BAE4E4D992C21 CRC64;

Query Match 60.7%; Score 34; DB 1; Length 264;

Best Local Similarity 62.5%; Pred. No. 18;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 BEVVPXGMHY 10

Db 235 WVPAGLHP 242

RESULT 12

SLSI_YARLI STANDARD; PRT; 426 AA.
 ID SLSI_YARLI
 AC Q99158;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE SLS1 protein precursor.
GN SLS1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]_SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 20460 / W29;
RX MEDLINE=96216076; PubMed=8662639;
RA Boissacq A., Beckerich J.-M., Gaillardin C.,
RT "Sls1p, an endoplasmic reticulum component, is involved in the
RT protein translocation process in the yeast Yarrowia lipolytica.";
RL J. Biol. Chem. 271:11668-11675(1996).
CC -!- FUNCTION: Involved in the protein translocation process. May
CC interact directly with translocating polypeptides to facilitate
CC their transfer and/or help their folding in the ER. It is not
CC required for viability but is essential for optimal growth at
CC elevated temperatures.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z50154; CAA90516.1; -.
DR PIR; S58132; S58132.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; 1.
KW Endoplasmic reticulum; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 426 SLS1 PROTEIN.
FT SITE 423 426 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 426 AA; 47201 MW; 0ACD7EF17540B82 CRC64;
Query Match 60.7%; Score 34; DB 1; Length 426;
Best Local Similarity 44.4%; Pred. No. 30;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMH 9
Db :|:|:|:|:
52 DQVIPAGLH 60
RESULT 13
YL15 VIBPA STANDARD; PRT; 441 AA.
ID YL15 VIBPA
AC P46231;
DT 01-NOV-1995 (Rel. 32, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein VP2115 (ORF3).
DE VP2115.
GN Vibrio parahaemolyticus.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Bacteriia; Proteobacteria; Vibrrio.
OX NCBI_TaxID=670;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Tajima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";

RL Lancet 361:743-749(2003).
RN [2]
RN SEQUENCE OF 1-140 FROM N.A.
RC STRAIN=BB22;
RX MEDLINE=94292449; PubMed=8021208;
RA McCarter L.L.;
RT "MecY, a component of the sodium-type flagellar motor.";
RL J. Bacteriol. 176:4219-4225(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI0325.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP005080; BAC60378.1; -.
DR EMBL; U06949; AAA21571.1; -.
DR InterPro; IPR004770; Antiport_nhaC.
DR InterPro; IPR001991; Na/H_antiporter; 1.
DR Pfam; PF03553; Na_H_antiporter; 1.
KW PRINTS; PR00173; EDRNSPORT.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 324 354 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 441 AA; 45961 MW; 451969FE307E4D46 CRC64;
Query Match 58.9%; Score 33; DB 1; Length 441;
Best Local Similarity 54.5%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGMHYS 11
Db :|:|:|:|:
417 ETVPFTFIHN 427
RESULT 14
PHSL DESBA STANDARD; PRT; 513 AA.
ID PHSL DESBA
AC P13065;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Periplasmic [Nifese] hydrogenase large subunit (BC 1.12.99.6) (Nifese
DE hydrogenlyase large chain).
OS Desulfovibrio baculatus (Desulfomicrobium baculatus).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfomicrobiaceae; Desulfomicrobium.
OX NCBI_TaxID=899;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88058744; PubMed=3316183;
RA Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;
RT "Cloning and sequencing of the genes encoding the large and small
RT subunits of the periplasmic (Nifese) hydrogenase of Desulfovibrio
RT baculatus.";
RL J. Bacteriol. 169:5401-5407(1987).
RN [2]
RN REVISIONS.
RP Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;

J. Bacteriol. 170:4429-4429 (1988).
 [3]
 X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 MEDLINE=99306038; PubMed=10378275;
 Garcin E., Vernece X., Hatchikian B.C., Volbeda A., Frey M.,
 Fontecilla-Camps J.C.;
 Removal of the bridging ligand atom at the Ni-Fe active site of
 [NiFe] hydrogenase upon reduction with H₂, as revealed by X-ray
 structure analysis at 1.4-A resolution.";
 Structure 7:557-566 (1999).
 CC -!- CATALYTIC ACTIVITY: H(2) + acceptor = 2 H(+) + reduced acceptor.
 CC -!- COFACTOR: Nickel, 2 irons and selenocysteine. Iron 1 has three
 CC cyanide and carbon monoxide ligands. Iron 2 has three water
 CC ligands.
 CC -!- SUBUNIT: Heterodimer of a large and a small subunit.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- MISCELLANEOUS: PERHAPS THE LEADER OF THE SMALL SUBUNIT SERVES AS A
 CC TRANSPORT VEHICLE FOR BOTH SUBUNITS.
 CC -!- SIMILARITY: Belongs to the [NiFe]/[NiFeSe] hydrogenase large
 CC subunit family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M18271; AAA23375.2; -.
 CC PIR: A33101; HQDVLB.
 CC PDB: 1CC1; 01-JUN-99.
 CC InterPro: IPR001501; Ni_hdl.
 CC Pfam: PF00374; Nifese_Hases; 1.
 CC PROSITE: PS00507; N1_HGENASE_L1; 1.
 CC PROSITE: PS00508; N1_HGENASE_L2; 1.
 CC Oxidoreductase; Periplasmic; Metal-binding; Nickel; Iron; Selenium;
 CC Selenocysteine; 3D-structure.
 CC INIT MET 0 0
 CC METAL 51 51 IRON 2.
 CC METAL 70 70 NICKEL.
 CC METAL 73 73 IRON 1.
 CC METAL 73 73 NICKEL.
 CC METAL 444 444 IRON 2 (VIA CARBONYL OXYGEN).
 CC METAL 492 492 NICKEL.
 CC METAL 495 495 IRON 1.
 CC METAL 495 495 NICKEL.
 CC METAL 498 498 IRON 2.
 CC SE CYS 492 492
 CC STRAND 13 16
 CC STRAND 23 23
 CC STRAND 26 33
 CC TURN 34 35
 CC STRAND 36 44
 CC STRAND 46 46
 CC HELIX 50 53
 CC TURN 54 56
 CC HELIX 59 61
 CC HELIX 62 65
 CC HELIX 66 69
 CC HELIX 74 89
 CC TURN 90 90
 CC HELIX 95 119
 CC TURN 120 121
 CC HELIX 122 124
 CC TURN 125 125
 CC TURN 132 133
 CC TURN 139 140
 CC HELIX 144 147
 CC HELIX 151 183
 CC STRAND 192 193
 CC TURN 194 195
 CC STRAND 196 197

FT HELIX 203 222
 FT TURN 223
 FT HELIX 224 234
 FT HELIX 236 239
 FT STRAND 248 250
 FT STRAND 253 255
 FT TURN 258 259
 FT STRAND 263 265
 FT STRAND 268 271
 FT TURN 272 273
 FT STRAND 274 276
 FT HELIX 280 282
 FT STRAND 283 286
 FT TURN 288 289
 FT STRAND 290 290
 FT TURN 291 291
 FT STRAND 292 292
 FT STRAND 300 301
 FT HELIX 302 304
 FT STRAND 309 309
 FT TURN 311 312
 FT TURN 314 315
 FT STRAND 318 318
 FT STRAND 322 325
 FT TURN 326 327
 FT STRAND 328 328
 FT STRAND 331 331
 FT HELIX 334 341
 FT STRAND 344 344
 FT HELIX 347 357
 FT STRAND 363 363
 FT HELIX 364 371
 FT HELIX 374 388
 FT TURN 401 402
 FT STRAND 406 406
 FT STRAND 415 423
 FT TURN 424 425
 FT STRAND 426 435
 FT TURN 436 437
 FT STRAND 438 446
 FT TURN 447 447
 FT HELIX 448 451
 FT TURN 457 458
 FT HELIX 463 468
 FT TURN 469 470
 FT STRAND 472 472
 FT TURN 476 477
 FT HELIX 480 488
 FT TURN 489 489
 FT STRAND 492 492
 FT HELIX 493 497
 SQ SEQUENCE 513 AA; 56683 MW; AC8285A6F80576FC CRC64;
 Query Match 58.9%; Score 33; DB 1; Length 513;
 Best Local Similarity 71.4%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 5 PGMHYS 11
 Db 297 PGLHYS 303
 RESULT 15
 MUTL_BACSU STANDARD; PRT; 627 AA.
 ID MUTL_BACSU
 AC P49850;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR BSU17050.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis mutL operon: identification, nucleotide
sequence and mutagenesis";
RL Microbiology 142:2021-2029(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriello R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Denzian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol K., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).

KW DNA repair; Complete proteome.
SQ SEQUENCE 627 AA; 70431 MW; 068A0509CC265343 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 627;
Best Local Similarity 54.5%; Pred.No. 71;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 REVVPXGMHYS 11
DB 488 EMIVELTTHYS 498

Search completed: June 3, 2004, 11:49:51
Job time : 5.86667 secs

CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U27343; AAB19236.1; -;
CC EMBL: Z99112; CAB13578.1; -;
CC PIR: A69663; A69663.
CC HSP: P23367; IBKN.
CC Subtilist; BG11402; mutL.
CC HAMAP: MF_00149; -; 1.
CC InterPro: IPR003594; ATPbind_ATPase.
CC InterPro: IPR002099; DNA_mis_repair.
CC Pfam: PF01119; DNA_mis_repair; 1.
CC Pfam: PF02518; HATPase_C; 1.
CC SMART: SM00387; HATPase_C; 1.
CC TIGRFAMs: TIGR00595; mutL; 1.
CC PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.

NO.	SCORE	MATCH	LENGTH	DS	ID	DESCRIPTION
1	41	73.2	413	11	Q8K289	mus musculus
2	39	69.6	1057	16	Q8CPJ4	staphylococ
3	37	66.1	308	16	Q9X2E2	thermotoga
4	37	66.1	322	17	Q9HLH8	thermoplas
5	37	66.1	1044	16	Q8DTH0	synechococ
6	36	64.3	208	2	Q46486	corynebacte
7	36	64.3	252	16	Q8YWP1	anabaena sp
8	36	64.3	819	10	Q9AVK4	pisum sativ
9	35	62.5	139	2	Q57489	bacteroides
10	35	62.5	156	16	Q9PC35	xyella fas
11	35	62.5	156	16	Q87D36	xyella fas
12	35	62.5	233	10	Q40479	niotiana t
13	35	62.5	237	10	Q9LW50	niotiana s
14	35	62.5	262	16	Q7WNB7	borderella
15	35	62.5	262	16	Q7W0Z3	borderella
16	35	62.5	317	9	Q38317	lactobacill

```

DE Carbamoyl-phosphate synthase large chain.
GN S20879
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE016746; AAC004476.1; -
DR GO: GO:0005737; Cytoplasm; IEA.
DR GO: GO:0005524; F-ATP binding; IEA.
DR GO: GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
DR GO: GO:0004197; F:cysteine-type dehydratase activity; IEA.
DR GO: GO:0016874; Filigase activity; IEA.
DR GO: GO:0006526; P:arginine biosynthesis; IEA.
DR GO: GO:0006807; P:nitrogen metabolism; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro: IPR006275; CarA_L_glu.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR005482; MGS_Like.
DR InterPro: IPR004362; SHprt_acsite.
DR Pfam: PF00289; CPase_L_chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS; PR00098; CPASE.
DR TIGRFAMs; TIGR01369; CPaseII_lg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR Complete proteome.
RW SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;
SQ
Query Match 69.6%; Score 39; DB 16; Length 1057;
Best Local Similarity 63.6%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 BEVVPXGMHYS 11
DB 189 KEVVSNGHYS 199
RP SEQUENCE FROM N.A.
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."

```

```

RL Nature 399:323-329(1999).
DR EMBL: AE001819; AAD36885.1; -.
DR TIGR: TW1822; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR001972; Stomat.
DR Pfam: PF01145; Band_7; 1.
DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
DR Protease; Complete proteome.
RW SEQUENCE 308 AA; 34778 MW; ADE03603E5101A9D CRC64;
SQ
Query Match 66.1%; Score 37; DB 16; Length 308;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMHY 10
DB 41 VVPSGIHY 48
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RA "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL: AL445063; CAC11395.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
DR Complete proteome.
RW SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;
SQ
Query Match 66.1%; Score 37; DB 17; Length 322;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMHY 10
DB 66 VVPDGLHY 73
RP SEQUENCE FROM N.A.
RC MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RA "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL: AL445063; CAC11395.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
DR Complete proteome.
RW SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;
SQ
Query Match 66.1%; Score 37; DB 17; Length 322;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMHY 10
DB 66 VVPDGLHY 73
RP SEQUENCE FROM N.A.
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."

```

NCBI_TaxID=32046;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohza M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AF005374; BAC09170.1; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001036; Acflvin_res.
 DR InterPro; IPR004764; HAS1.
 DR Pfam; PF00873; ACB_tran.1.
 DR PRINTS; PRO0702; ACRIFLAVINRP.
 DR TIGRFAMs; TIGR00915; 2A0602; 1.
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
 Query Match 66.1%; Score 37; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EVVVPXGMHYS 11
 DB 843 EVVLENGIGYS 853
 RESULT 6
 ID Q46486 PRELIMINARY; PRT; 208 AA.
 AC Q46486;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Hypothetical protein (Gcra).
 GN GCRA.
 OS Corynebacterium xerosis, and
 OS Corynebacterium striatum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1725; 43770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=C.xerosis; STRAIN=M82B;
 RX MEDLINE=96117603; PubMed=8559800;
 RA Tauch A., Kassing F., Kalinowski J., Puhler A.;
 RT "The Corynebacterium xerosis composite transposon Tn5432 consists of
 RT two identical insertion sequences, designated ISI249, flanking the
 RT erythromycin resistance gene ermCX.";
 RL Plasmid 34:119-131(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=C.striatum; STRAIN=M82B;
 RX MEDLINE=20194806; PubMed=1073268;
 RA Tauch A., Krieff S., Kalinowski J., Puhler A.;
 RT isolate Corynebacterium pp10 from the multiresistant clinical
 RT initially identified in soil bacteria and in plant, animal, and human
 RT pathogens";
 RL Mol. Gen. Genet. 263:1-11(2000).
 DR EMBL; U21300; AAC95478.1; --
 DR EMBL; AF024666; AAG03390.1; --
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 208 AA; 23012 MW; F1504BE1EFCDB85A6 CRC64;
 Query Match 64.3%; Score 36; DB 2; Length 208;

Best Local Similarity 50.0%; Pred. No. 36;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVVPXGMHYS 11
 DB 130 DVIPEGKHYA 139
 RESULT 7
 ID Q8YWP1 PRELIMINARY; PRT; 252 AA.
 AC Q8YWP1;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Hypothetical protein Alr1563.
 GN ALR1563.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003586; BAB77929.1; --
 DR FIR; AE2001; AE2001.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 252 AA; 28831 MW; 925572DA5D1CA519 CRC64;
 Query Match 64.3%; Score 36; DB 16; Length 252;
 Best Local Similarity 50.0%; Pred. No. 44;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EVVVPXGMHY 10
 DB 235 EMIVPAGLHF 244
 RESULT 8
 ID Q9AVK4 PRELIMINARY; PRT; 819 AA.
 AC Q9AVK4;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE SCARECROW.
 GN PSSCR.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Alaska;
 RX MEDLINE=21231727; PubMed=11333309;
 RA Sassa N., Matsushita Y., Nakamura T., Nyunoya H.;
 RT "The Molecular Characterization and in situ Expression Pattern of Pea
 RT SCARECROW Gene";
 RL Plant Cell Physiol. 42:385-394(2001).
 DR EMBL; AB048713; BAB39155.1; --
 DR TRANSFAC; T05513; --
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.

```
DR InterPro; IPR001444; Flag_bb_rod.
DR InterPro; IPR005202; GRAS.
DR Pfam; PF03514; GRAS; 1
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1
SQ SEQUENCE 819 AA; 90372 MW; 418678D6D72ADFA CRC64;

Query Match 64.3%; Score 36; DB 10; Length 819;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
Db 343 DDVPTSLHPS 353
:::|:|:|
:::|:|:|

RESULT 9
ID Q57489 PRELIMINARY; PRT; 139 AA.
AC Q57489;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DNA ligase (Fragment).
OS Bacteroides nodosus (Dichelobacter nodosus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Cardiobacteriales;
OC Cardiobacteriaceae; Dichelobacter.
OX NCBI_TaxID=870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96020672; PubMed=7476204;
RA Moses E.K., Good R.T., Sinistaj M., Billington S.J., Langford C.J.,
RA Rood J.I.;
RT "A multiple site-specific DNA-inversion model for the control of Omp1
RT phase and antigenic variation in Dichelobacter nodosus.";
RL Mol. Microbiol. 17:183-196(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96257263; PubMed=8654969;
RA Billington S.J., Sinistaj M., Cheetham B.F., Ayres A., Moses E.K.,
RA Katz M.E., Rood J.I.;
RT "Identification of a native Dichelobacter nodosus plasmid and
RT implications for the evolution of the vap regions.";
RL Gene 172:111-116(1996).
DR EMBL; U02462; AAB12366.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 1.
DR SMART; SM00292; BRCT; 1.
DR PROSITE; PS00172; BRCT; 1.
KW Ligase.
FT NON_TER
FT TER
SQ SEQUENCE 139 AA; 15052 MW; E0E110AA4B7D4708 CRC64;

Query Match 62.5%; Score 35; DB 2; Length 139;
Best Local Similarity 55.6%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHYS 11
Db 21 IVPAGVHNS 29
:::|:|:|
:::|:|:|

RESULT 10
Q9PC35 PRELIMINARY; PRT; 156 AA.
AC Q9PC35;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein Xf1950.
GN Xf1950.
OS Xylella fastidiosa.

Query Match 62.5%; Score 35; DB 16; Length 156;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
Db 119 BEILPQGVH 127
:::|:|:|
:::|:|:|

RESULT 11
Q87D36 PRELIMINARY; PRT; 156 AA.
AC Q87D36;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN PD0850.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=945C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Frega J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silvestri W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004014; AAF84752.1; -.
DR FIR; D82618; D82618.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007155; P:signal transduction; IEA.
DR InterPro; IPR002545; Chew.
DR PROSITE; PS00851; CHEW; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 156 AA; 17144 MW; D8358619C6671A5D CRC64;

Query Match 62.5%; Score 35; DB 16; Length 156;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
Db 119 BEILPQGVH 127
:::|:|:|
:::|:|:|

RESULT 11
Q87D36 PRELIMINARY; PRT; 156 AA.
AC Q87D36;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN PD0850.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
```

RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Teai S.M.,
 RA Carreir H., Carreiro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
 RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
 RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
 RA Baia G.S., Blanco S.R., Brito M.S., Camnayan F.S., Celestino A.V.,
 RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
 RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
 RA Civerolo E.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
 RA Kitajima J.P.;
 RT "Comparative analyses of the complete genome sequences of Pierce's
 RT disease and citrus variegated chlorosis strains of Xylella
 RT fastidiosa";
 RL J. Bacteriol. 185:1018-1026(2003).
 DR EMBL; AB012556; AO028718.1; -.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0004871; F:signal transducer activity; IEA.
 DR GO; GO:0006935; P:chemotaxis; IEA.
 DR GO; GO:0007165; P:signal transduction; IEA.
 DR InterPro: IPR002545; ChEM.
 DR PROSITE; PS50851; CHEW; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 156 AA; 17130 MW; D83583B9C6671A5D CRC64;
 Query Match 62.5%; Score 35; DB 16; Length 156;
 Best Local Similarity 55.6%; Pred. No. 42;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHY 9
 DB 119 EEILPQGVH 127
 RESULT 12
 ID Q40479 PRELIMINARY; PRT; 233 AA.
 AC Q40479;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE EREP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC lamids; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4097;
 RN [1]
 SQ SEQUENCE FROM N.A.
 RC STRAIN-BY4; TISSUE=Leaf;
 RX MEDLINE=95276459; PubMed=7756828;
 RA Chme-Takagi M., Shinshi H.;
 RT "Ethylene-inducible DNA binding proteins that interact with an
 RT ethylene-responsive element";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38126; BAA07324.1; -.
 DR PIR; T02590; T02590.
 DR HSSP; O80337; 2GCC.
 DR TRANSFAC; T02654; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETRSPPELMNT.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;
 Query Match 62.5%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
 DB 90 QAVVPKGRHY 99
 RESULT 13
 ID Q9LW50 PRELIMINARY; PRT; 237 AA.
 AC Q9LW50;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Ethylene-responsive element binding factor.
 OS Nicotiana sylvestris (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC lamids; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4096;
 RN [1]
 SQ SEQUENCE FROM N.A.
 RC MEDLINE=20399450; PubMed=10945353;
 RA Kitajima S., Kovama T., Ohme-Takagi M., Shinshi H., Sato F.;
 RT "Characterization of gene expression of NERFs, transcription factors
 RT of basic PR genes from Nicotiana sylvestris";
 RL Plant Cell Physiol 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1; -.
 DR HSSP; O80337; 2GCC.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETRSPPELMNT.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 237 AA; 26243 MW; 01BC3EEB51E46298 CRC64;
 Query Match 62.5%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHY 10
 DB 94 QAVVPKGRHY 103
 RESULT 14
 ID Q7WNB7 PRELIMINARY; PRT; 262 AA.
 AC Q7WNB7;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative enoyl-CoA hydratase.
 GN B01123.
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 NCBI_TaxID=518;
 RN [1]
 SQ SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebahia M., Preston A., Churcher C.M., Mungall K.L.,
 RA Harris D.E., Holden M.T.G., Churche K., Harris B., Quail M.A.,
 RA Cerdano-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagers K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640440; CAE31621.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28907 MW; B3CA29331CB776B2 CRC64;

Query Match 62.5%; Score 35; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 9
:|||||
Db 182 QEVVPYGOH 190

RESULT 15

Q7W0Z3 PRELIMINARY; FRT; 262 AA.
AC Q7W0Z3;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BPP0913.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-597;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640445; CAE40321.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28888 MW; 31CA2935EBB776BD CRC64;

Query Match 62.5%; Score 35; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 9
:|||||
Db 182 QEVVPYGOH 190

Search completed: June 3, 2004, 11:57:31
Job time : 30.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.684 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 BEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_29Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	11	5 ABB80561	Abb80561 Hepatitis
2	52	96.3	11	5 ABB80524	Abb80524 Hepatitis
3	52	96.3	11	5 ABB80529	Abb80529 Hepatitis
4	52	96.3	11	5 ABB80528	Abb80528 Hepatitis
5	52	96.3	11	5 ABB80562	Abb80562 Hepatitis
6	47	87.0	11	5 ABB80542	Abb80542 Hepatitis
7	47	87.0	11	5 ABB80543	Abb80543 Hepatitis
8	47	87.0	11	5 ABB80538	Abb80538 Hepatitis
9	46	85.2	11	5 ABB80525	Abb80525 Hepatitis
10	46	85.2	11	5 ABB80548	Abb80548 Hepatitis
11	46	85.2	11	5 ABB80521	Abb80521 Hepatitis
12	46	85.2	11	5 ABB80522	Abb80522 Hepatitis
13	46	85.2	11	5 ABB80547	Abb80547 Hepatitis
14	46	85.2	11	5 ABB80566	Abb80566 Hepatitis
15	46	85.2	11	5 ABB80556	Abb80556 Hepatitis
16	46	85.2	11	5 ABB80557	Abb80557 Hepatitis
17	46	85.2	11	5 ABB80551	Abb80551 Hepatitis
18	46	85.2	11	5 ABB80563	Abb80563 Hepatitis
19	46	85.2	11	5 ABB80565	Abb80565 Hepatitis
20	46	85.2	11	5 ABB80567	Abb80567 Hepatitis
21	46	85.2	11	5 ABB80559	Abb80559 Hepatitis
22	46	85.2	11	5 ABB80526	Abb80526 Hepatitis
23	46	85.2	11	5 ABB80564	Abb80564 Hepatitis
24	46	85.2	11	5 ABB80568	Abb80568 Hepatitis
25	45	83.3	11	5 ABB80523	Abb80523 Hepatitis

26	45	83.3	11	5 ABB80558	Abb80558 Hepatitis
27	45	83.3	11	5 ABB80560	Abb80560 Hepatitis
28	45	83.3	11	5 ABB80527	Abb80527 Hepatitis
29	44	81.5	11	5 ABB80534	Abb80534 Hepatitis
30	44	81.5	11	5 ABB80533	Abb80533 Hepatitis
31	41	75.9	11	5 ABB80536	Abb80536 Hepatitis
32	41	75.9	11	5 ABB80535	Abb80535 Hepatitis
33	41	75.9	11	5 ABB80540	Abb80540 Hepatitis
34	41	75.9	11	5 ABB80539	Abb80539 Hepatitis
35	40	74.1	11	5 ABB80549	Abb80549 Hepatitis
36	40	74.1	11	5 ABB80544	Abb80544 Hepatitis
37	40	74.1	11	5 ABB80537	Abb80537 Hepatitis
38	40	74.1	11	5 ABB80541	Abb80541 Hepatitis
39	40	74.1	11	5 ABB80553	Abb80553 Hepatitis
40	40	74.1	11	5 ABB80552	Abb80552 Hepatitis
41	40	74.1	11	5 ABB80545	Abb80545 Hepatitis
42	39	72.2	11	5 ABB80546	Abb80546 Hepatitis
43	39	72.2	11	5 ABB80554	Abb80554 Hepatitis
44	39	72.2	11	5 ABB80550	Abb80550 Hepatitis
45	39	72.2	11	5 ABB80555	Abb80555 Hepatitis

ALIGNMENTS

RESULT 1					
ABB80561					
ID	ABB80561	standard; peptide; 11 AA.			
XX	AC	ABB80561;			
XX	DT	08-OCT-2002 (first entry)			
XX	DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.			
XX	XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketocamide; vitucide.			
XX	OS	Synthetic.			
XX	XX	Key	Location/Qualifiers		
FT	Key	Modified-site	1	/note= "N-terminal acetyl"	
FT	Modified-site	6	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"		
FT	Modified-site	8	/note= "Oxymethionine"		
FT	Misc-difference	8	/note= "D-form residue"		
FT	Modified-site	11	/note= "C-terminal amide"		
WO200208251-A2.					
31-JAN-2002.					
19-JUL-2001; 2001WO-US023169.					
21-JUL-2000; 2000US-0220101P.					
(CORV-) CORVAS INT INC.					
Lim-Wilby M, Levy OE, Brunck TK;					
WPI; 2002-361643/39.					
Novel peptide compound having hepatitis C virus protease inhibitory					
PT activity useful for treating disorders associated with hepatitis C virus					
protease.					
Claim 17; Page 65; 69pp; English.					

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 2
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
AC ABB80524;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 9 residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX

SQ Sequence 11 AA;
Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 3
ABB80529
ID ABB80529 standard; peptide; 11 AA.
XX
AC ABB80529;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 9 residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
 XX
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX
 KW virucide.
 XX
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGQDYS 11
 RESULT 7
 ID ABB80543 standard; peptide; 11 AA.
 AC ABB80543;
 XX
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 OS

PH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGQDYS 11
 RESULT 8
 ID ABB80538 standard; peptide; 11 AA.
 AC ABB80538;
 XX
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS
 PH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT

XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Sequence 11 AA;
PS
XX
XX Query Match 85.2%; Score 46; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.036;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 12
ABB80521
ID ABB80521 standard; peptide; 11 AA.
XX
XX ABB80521;
AC
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Sequence 11 AA;
PS
XX
XX Query Match 85.2%; Score 46; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.036;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 12
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
XX ABB80522;
AC
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Sequence 11 AA;
SQ

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMSYS 11

RESULT 13
ABB80547
ID ABB80547 standard; peptide; 11 AA.
XX AC ABB80547;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX PN 31-JAN-2002.
XX PD 19-JUL-2001; 2001WO-US023169.
XX PF 21-JUL-2000; 2000US-0220101P.
XX PR (CORV-) CORVAS INT INC.
XX PA Lim-Wilby M, Levy OE, Brunck TK;
XX PI WPI; 2002-361643/39.
XX DR Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGTDYS 11

RESULT 14
ABB80566
ID ABB80566 standard; peptide; 11 AA.
XX AC ABB80566;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "2-aminoisobutyl carbonyl residue forming a keto
FT -amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX PN 31-JAN-2002.
XX PD 19-JUL-2001; 2001WO-US023169.
XX PF 21-JUL-2000; 2000US-0220101P.
XX PR (CORV-) CORVAS INT INC.
XX PA Lim-Wilby M, Levy OE, Brunck TK;
XX PI WPI; 2002-361643/39.
XX DR Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMSYS 11

RESULT 15
ABB80556
ID ABB80556 standard; peptide; 11 AA.
XX AC ABB80556;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #36.
XX

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
PI WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
| | | | | | | | | |
Db 1 BEVVPXGSDYS 11

Search completed: June 3, 2004, 11:48:23
Job time : 46.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-8
Perfect score: 54
Sequence: 1 BEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	66.7	181	4	US-09-134-000C-4848
2	36	66.7	341	3	US-08-853-948B-4
3	36	66.7	348	3	US-08-853-948B-5
4	36	66.7	368	4	US-09-697-367-24
5	36	66.7	1049	4	US-08-394-272-10
6	36	66.7	1068	2	US-08-429-054A-11
7	36	66.7	1068	2	US-08-718-777-7
8	36	66.7	1068	3	US-09-051-341-7
9	36	66.7	1068	4	US-09-394-272-8
10	36	66.7	1081	4	US-09-394-272-4
11	36	66.7	1083	4	US-09-394-272-11
12	36	66.7	1084	4	US-09-394-272-9
13	34	63.0	140	3	US-08-569-147-76
14	34	63.0	140	3	US-08-569-147-82
15	34	63.0	1065	4	US-09-252-991A-31637
16	34	63.0	3472	4	US-09-408-020-4
17	33	61.1	59	4	US-08-963-851-14
18	33	61.1	237	4	US-09-540-236-2677
19	33	61.1	303	4	US-09-134-000C-4318
20	33	61.1	378	1	US-08-070-165F-8
21	33	61.1	378	2	US-08-885-418-8
22	33	61.1	473	4	US-09-252-991A-26805
23	33	61.1	765	4	US-09-819-989-4
24	33	61.1	765	4	US-10-273-952-4
25	33	61.1	801	3	US-08-383-630-6
26	33	61.1	811	3	US-09-819-989-2
27	33	61.1	811	4	US-10-273-952-2

28 33 61.1 883 4 US-09-667-373-4 Sequence 4, Appli
29 32 59.3 65 6 517197-51 Patent No. 517197
30 32 59.3 102 2 US-08-580-988A-23 Sequence 23, Appli
31 32 59.3 152 2 US-08-460-694-4 Sequence 4, Appli
32 32 59.3 152 3 US-08-460-744-4 Sequence 4, Appli
33 32 59.3 152 3 US-07-657-711B-4 Sequence 4, Appli
34 32 59.3 173 1 US-08-193-977-7 Sequence 7, Appli
35 32 59.3 189 2 US-08-464-517-21 Sequence 21, Appli
36 32 59.3 189 2 US-08-246-361A-21 Sequence 21, Appli
37 32 59.3 189 3 US-08-463-772-21 Sequence 21, Appli
38 32 59.3 189 5 PCT-US93-05000-21 Sequence 21, Appli
39 32 59.3 231 3 US-08-926-842B-20 Sequence 20, Appli
40 32 59.3 236 2 US-08-464-517-22 Sequence 22, Appli
41 32 59.3 236 2 US-08-246-361A-22 Sequence 22, Appli
42 32 59.3 236 3 US-08-463-772-22 Sequence 22, Appli
43 32 59.3 236 5 PCT-US93-05000-22 Sequence 22, Appli
44 32 59.3 240 3 US-08-926-842B-21 Sequence 21, Appli
45 32 59.3 241 4 US-09-107-532A-4172 Sequence 4172, Ap

ALIGNMENTS

RESULT 1
US-09-134-000C-4848
; Sequence 4848, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Staun et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4848
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 66.7%; Score 36; DB 4; Length 181;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10
|||
DB 145 BEVVPXGMDY 154

RESULT 2
US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURE:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe,

; OTHER INFORMATION: Pro, Ser, Thr, Trp, Tyr, or Val
US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 228 VPPGMDFS 236

RESULT 3

US-08-853-948B-5

; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 234 VPPGMDFS 242

RESULT 4

US-09-697-367-24

; Sequence 24, Application US/09697367
; Patent No. 6323015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Jarczyński, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BB1166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 217 VPPGMDFS 225

RESULT 5

US-09-394-272-10

; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 436 VPPGMDFS 444

RESULT 6

US-08-429-054A-11

; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN, VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHETASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/429,054A
FILING DATE: 26-APR-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 842,337
FILING DATE: 20-March-1992
APPLICATION NUMBER: PCT/FR 91/00593
FILING DATE: 18-July-1991
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: French 90402094.9
FILING DATE: 20-July-1990
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Charles A. Musierlian
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146.1137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000

```
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
US-08-429-054A-11
Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPPGMDFS 443

RESULT 7
US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE
; TITLE OF INVENTION: PHOSPHATE
; TITLE OF INVENTION: SYNTHASE IN PLANTS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Barbara Rae-Venter
; STREET: 260 Sheridan Avenue, Suite 440
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,777
; FILING DATE: NOT YET ASSIGNED
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,471
; FILING DATE: 27-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara Rae-Venter
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE.072.02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)328-4400
; TELEFAX: (415)328-4477
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-777-7
Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPPGMDFS 443

RESULT 8
US-09-051-341-7
; Sequence 7, Application US/09051341
; Patent No. 6124528
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, C. K.
; TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rae-Venter Law Group, P.C.
; STREET: 260 Sheridan Avenue, Suite 440
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,341
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17351
; FILING DATE: 25-OCT-1996
; APPLICATION NUMBER: US 08/549,016
; FILING DATE: 27-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/372,200
; FILING DATE: 12-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara Rae-Venter, Ph.D.,
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE.110.02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)328-4400
; TELEFAX: (415)328-4477
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-051-341-7
Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPPGMDFS 443

RESULT 9
US-09-394-272-8
; Sequence 8, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
```

us-09-909-164-8.ra1

Fri Jun 4 16:16:20 2004

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 1068

; TYPE: PRT

; ORGANISM: Zea mays

US-09-394-272-8

Query Match 66.7%; Score 36; DB 4; Length 1068;

Best Local Similarity 66.7%; Pred. No. 81;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

|:|:|:|:|

DB 435 VIPPGMDFS 443

RESULT 10

US-09-394-272-4

; Sequence 4, Application US/09394272

; Patent No. 6472588

; GENERAL INFORMATION:

; APPLICANT: Haigler, Candace H.

; APPLICANT: Holaday, A. Scott

; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED

; FILE REFERENCE: 201304/1000

; CURRENT APPLICATION NUMBER: US/09/394,272

; CURRENT FILING DATE: 1999-09-10

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1081

; TYPE: PRT

; ORGANISM: Craterostigma plantagineum

US-09-394-272-4

Query Match 66.7%; Score 36; DB 4; Length 1081;

Best Local Similarity 66.7%; Pred. No. 83;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

|:|:|:|:|

DB 445 VIPPGMDFS 453

RESULT 11

US-09-394-272-11

; Sequence 11, Application US/09394272

; Patent No. 6472588

; GENERAL INFORMATION:

; APPLICANT: Haigler, Candace H.

; APPLICANT: Holaday, A. Scott

; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED

; FILE REFERENCE: 201304/1000

; CURRENT APPLICATION NUMBER: US/09/394,272

; CURRENT FILING DATE: 1999-09-10

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11

; LENGTH: 1083

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-394-272-11

Query Match 66.7%; Score 36; DB 4; Length 1083;

Best Local Similarity 66.7%; Pred. No. 83;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

|:|:|:|:|

DB 483 VIPPGMDFS 491

RESULT 12

US-09-394-272-9

; Sequence 9, Application US/09394272

; Patent No. 6472588

; GENERAL INFORMATION:

; APPLICANT: Haigler, Candace H.

; APPLICANT: Holaday, A. Scott

; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED

; FILE REFERENCE: 201304/1000

; CURRENT APPLICATION NUMBER: US/09/394,272

; CURRENT FILING DATE: 1999-09-10

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1084

; TYPE: PRT

; ORGANISM: Oryza sativa

US-09-394-272-9

Query Match 66.7%; Score 36; DB 4; Length 1084;

Best Local Similarity 66.7%; Pred. No. 83;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

|:|:|:|:|

DB 453 VIPPGMDFS 461

RESULT 13

US-08-569-147-76

; Sequence 76, Application US/08569147

; Patent No. 6180377

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: HUMANISED ANTIBODIES

; NUMBER OF SEQUENCES: 95

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

; ADDRESSEE: No. 6180377ris, LLP

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: U.S.A.

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/569,147

; FILING DATE: 25-March-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Trujillo, Doreen Yanko

; REGISTRATION NUMBER: 35,719

; REFERENCE/DOCKET NUMBER: CARP-0047

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 76:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 140 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-569-147-76

Query Match 63.0%; Score 34; DB 3; Length 140;

Best Local Similarity 75.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      3 VVFXGMDY 10
      ||| |||
Db      122 VVPTGFDY 129

RESULT 14
US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377is, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

      Query Match      63.0%; Score 34; DB 3; Length 140;
      Best Local Similarity 75.0%; Pred. No. 20;
      Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVFXGMDY 10
      ||| |||
Db      122 VVPTGFDY 129

RESULT 15
US-09-252-991A-31637
; Sequence 31637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31637
; LENGTH: 1065

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31637

      Query Match      63.0%; Score 34; DB 4; Length 1065;
      Best Local Similarity 85.7%; Pred. No. 2e+02;
      Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 PXGMDYS 11
      | ||| |
Db      324 PQGMDYS 330

Search completed: June 3, 2004, 12:03:07
Job time : 12.8 secs
```


OTHER INFORMATION: AMIDATION
US-09-909-164-8

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||
Db 1 EVVVPXGMDYS 11

RESULT 2

US-09-909-164-12
Sequence 12, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-13

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||
Db 1 EVVVPXGMDYS 11

RESULT 3

US-09-909-164-13
Sequence 13, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||
Db 1 EVVVPXGMDYS 11

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-13

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||
Db 1 EVVVPXGMDYS 11

RESULT 4

US-09-909-164-22
Sequence 22, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||
Db 1 EVVVPXGMDYS 11


```

; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-22

```

```

Query Match      87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 EEVVPXGMDYS 11
   |||||
Db 1 EEVVPXGQDYS 11

```

RESULT 5

```

US-09-909-164-26
; Sequence 26, Application US/09909164
; Publication No. US20020068702A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence

```

```

; OTHER INFORMATION: 11-mer synthesized according to example 1

```

```

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-26

```

```

Query Match      87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 EEVVPXGMDYS 11
   |||||
Db 1 EEVVPXGQDYS 11

```

RESULT 6

```

US-09-909-164-27

```

```

; Sequence 27, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita

```

```

; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence

```

```

; OTHER INFORMATION: 11-mer synthesized according to example 1

```

```

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-27

```

```

Query Match      87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 EEVVPXGMDYS 11
   |||||
Db 1 EEVVPXGQDYS 11

```

RESULT 7

```

US-09-909-164-5

```

```

; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K

```

```

; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence

```

```

; OTHER INFORMATION: 11-mer synthesized according to example 1

```

```

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)

```

OTHER INFORMATION: norvaline-(CO)

FEATURE: Brunk, Terence K

NAME/KEY: MOD RES

LOCATION: (1)..(11)

OTHER INFORMATION: AMIDATION

US-09-909-164-5

Query Match

Best Local Similarity 85.2%; Score 46; DB 12; Length 11;

Pred. No. 0.015;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11

|||||

Db 1 EVVVPXGMSYS 11

RESULT 8

US-09-909-164-6

Sequence 6, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunk, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

PRIOR FILING DATE: 2003-03-25

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

NAME/KEY: MOD RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (9)..(9)

OTHER INFORMATION: D-amino acid

FEATURE:

NAME/KEY: MOD RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

US-09-909-164-6

Query Match

Best Local Similarity 85.2%; Score 46; DB 12; Length 11;

Pred. No. 0.015;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11

|||||

Db 1 EVVVPXGMSYS 11

RESULT 9

US-09-909-164-9

Sequence 9, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunk, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

CURRENT FILING DATE: 2003-03-25

PRIOR APPLICATION NUMBER: 60/220,101

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 9

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

NAME/KEY: MOD RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (8)..(8)

OTHER INFORMATION: D-amino acid

US-09-909-164-9

Query Match

Best Local Similarity 85.2%; Score 46; DB 12; Length 11;

Pred. No. 0.015;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11

|||||

Db 1 EVVVPXGMSYS 11

RESULT 10

US-09-909-164-10

Sequence 10, Application US/09909164

Publication No. US20020068702A1

GENERAL INFORMATION:

APPLICANT: Corvas International, Inc.

APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E

APPLICANT: Brunk, Terence K

TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164

CURRENT FILING DATE: 2003-03-25

PRIOR APPLICATION NUMBER: 60/220,101

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn version 3.1

SEQ ID NO 10

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

NAME/KEY: MOD RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD RES

LOCATION: (11)..(11)

```

; APPLICANT: Levy, Odile E
; APELLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 32
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
;
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
;
US-09-909-164-32
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
        |||||
DB       1 EEVVPXGTDYS 11
        |||||

RESULT 13
US-09-909-164-35
; Sequence 35, Application US/09909164
; Publication NO. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 35
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
;
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
;
US-09-909-164-31
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
        |||||
DB       1 EEVVPXGMSYS 11
        |||||

RESULT 12
US-09-909-164-32
; Sequence 32, Application US/09909164
; Publication NO. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
;
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
;
US-09-909-164-31
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
        |||||
DB       1 EEVVPXGTDYS 11
        |||||

RESULT 11
US-09-909-164-31
; Sequence 31, Application US/09909164
; Publication NO. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
;
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
;
US-09-909-164-10
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
        |||||
DB       1 EEVVPXGMSYS 11
        |||||

```

us-09-909-164-8.rapb

Fri Jun 4 16:16:20 2004

```

; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)-(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-35

Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
   ||||| |||
Db 1 EEVVPXGSDYS 11

RESULT 14
US-09-909-164-40
; Sequence 40, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)-(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)-(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-41

Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
   ||||| |||
Db 1 EEVVPXGSDYS 11

Search completed: June 3, 2004, 12:57:15
Job time : 33.7333 secs

```

```

; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)-(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-35

Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
   ||||| |||
Db 1 EEVVPXGSDYS 11

RESULT 15
US-09-909-164-41
; Sequence 41, Application US/09909164

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EGVVPGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: Pirl.*

2: Pirl2.*

3: Pirl3.*

4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	2 S54619	hypothetical prote
2	38	70.4	363	2 D69551	conserved hypothet
3	36	66.7	102	2 A42452	VI protein - tobac
4	36	66.7	341	2 S72649	sucrose-phosphate
5	36	66.7	348	2 S72650	sucrose-phosphate
6	36	66.7	450	2 G96764	unknown protein P2
7	36	66.7	1049	2 JC4783	sucrose-phosphate
8	36	66.7	1068	1 JQ1329	sucrose-phosphate
9	36	66.7	1081	2 T09837	sucrose-phosphate
10	36	66.7	1083	2 T04062	sucrose-phosphate
11	36	66.7	1084	2 T04103	sucrose-phosphate
12	35	64.8	425	2 T24111	hypothetical prote
13	35	64.8	433	2 H87650	peptidoglycan-bind
14	35	64.8	440	2 H72784	probable alkaline
15	35	64.8	1150	2 T20173	hypothetical prote
16	35	64.8	1474	2 F69009	probable membrane
17	35	64.8	2747	2 B49132	fat facets (faf) s
18	34	63.0	99	2 S00210	plastocyanin b - L
19	34	63.0	155	2 S39255	plastocyanin precu
20	34	63.0	168	2 S58208	6-O-methylguanine-D
21	34	63.0	290	2 AG3104	6-O-methylguanine-D
22	34	63.0	290	2 D98182	hypothetical prote
23	34	63.0	296	2 F72745	probable hexosyltr
24	34	63.0	357	1 G69390	L-lactate dehydrog
25	34	63.0	366	2 G69350	ABC transporter AT
26	34	63.0	565	2 B36665	phenylalanine-tRNA
27	34	63.0	566	2 A70164	succinate dehydrog
28	34	63.0	587	2 F81138	disease resistance
29	34	63.0	906	2 T48998	

30	34	63.0	908	2 T48999	disease resistance
31	34	63.0	1062	2 F83335	RND multidrug effl
32	34	63.0	1062	2 T30830	hypothetical prote
33	34	63.0	3472	2 T31308	hypothetical 367K
34	33	61.1	97	2 A99427	partial transposas
35	33	61.1	128	2 A90471	hypothetical prote
36	33	61.1	172	2 S27021	fibroblast growth
37	33	61.1	184	2 E90335	hypothetical prote
38	33	61.1	225	2 S57810	hypothetical prote
39	33	61.1	257	2 A95001	conserved hypothet
40	33	61.1	257	2 A95446	unknown protein li
41	33	61.1	262	2 F90298	transposase ISC105
42	33	61.1	267	2 C90307	transposase ISC105
43	33	61.1	276	2 C64417	hypothetical prote
44	33	61.1	283	2 G83055	pantoate-beta-alan
45	33	61.1	299	2 E90487	transposase ISC105

ALIGNMENTS

RESULT 1

S54619

hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3

C/Species: Saccharomyces cerevisiae

C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C/Accession: S54619; S66879

R/de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A/Reference number: S54619

A/Accession: S54619

A/Molecule type: DNA

A/Residues: 1-156 <DEH>

A/Cross-references: EMBL:X87331; NID:g1041652; PIDN:CAA60762.1; PID:g829123

R/de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A/Reference number: S66877

A/Accession: S66879

A/Molecule type: DNA

A/Residues: 1-156 <DEW>

A/Cross-references: EMBL:Z74920; NID:g1420109; PIDN:CAA99201.1; PID:g1420111; MIPS:YORO

A/Experimental source: strain S288C

C/Genetics:

A/Cross-references: SGD:S0005539

A/Map position: 15R

C/Superfamily: hypothetical protein YOR013w

Query Match	74.1%	Score 40;	DB 2;	Length 156;
Best Local Similarity	77.8%	Pred. No. 1;		
Matches	7;	Conservative	1;	Mismatches
				Indels
				0;
				Gaps
				0;

Qy	2	EVVPGMDY 10
		:
Db	50	EVVPLGMDY 58

RESULT 2

D69551

conserved hypothetical protein Af2411 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C/Accession: D69551

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso

.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, B.F

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S

Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: D69551

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

RESULT 5

RESULT

JC4783
sucrose-phosphate synthase (EC 2.4.1.14) - rice
C:Species: *Oryza sativa* (rice)

C>Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
 C/Accession: J04783
 C/Valdez-Alarcon, J.J.; Ferrando, M.; Salerno, G.; Jimenez-Moraila, B.; Herrera-Estrella
 C/Date: 170, 217-222, 1996
 A/Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
 A/Reference number: J04783; MUID:96235138; PMID:8666248
 A/Molecule type: mRNA
 A/Accession: J04783
 A/Residues: 1-1049 <VAL>
 A/Cross-references: GB:U33175; NID:g1449931; PIDN:AAC49379.1; PID:g988270
 A/Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrose-phosphate-UDPglucosyl
 C/Comment: This enzyme catalyzes the formation of sucrose-phosphate form UDP-glucose and
 C/Genetics:
 A/Gene: Sps1
 A/Introns: 24/1, 103/3; 183/3; 205/3; 435/3; 519/3; 578/3; 596/3; 617/3; 931/3; 9
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: Glucosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:180-663/Domain: sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |||
 Db 436 VIPPGMDFS 444

RESULT 8
 J01329
 sucrose-phosphate synthase (EC 2.4.1.14) - maize
 C/Species: Zea mays (maize)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: J01329; PQ0260
 R/Worrell, A.C.; Bruneau, J.M.; Summerfelt, K.; Boersig, M.; Voelker, T.A.
 Plant Cell 3, 1121-1130, 1991
 A/Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
 A/Reference number: J01329; MUID:92338837; PMID:1840396
 A/Accession: J01329
 A/Molecule type: mRNA
 A/Residues: 1-1068 <WOR>
 A/Cross-references: GB:M97550; NID:g168625; PIDN:AAA33513.1; PID:g168626
 A/Accession: PQ0260
 A/Molecule type: protein
 A/Residues: 71-74; 206-212; 471-481; 872-892 <WOR1>
 C/Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
 C/Comment: This enzyme is involved in the regulation of carbon partitioning in the leave
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: Glucosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:178-666/Domain: sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |||
 Db 435 VIPPGMDFS 443

RESULT 9
 T09837
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantagineum
 C/Species: Craterostigma plantagineum
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T09837
 R/Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

Plant Physiol. 115, 113-121, 1997
 A/Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to sug
 A/Reference number: Z16874; MUID:97451773; PMID:9306694
 A/Accession: T09837
 A/Status: preliminary; translated from GB/EMBL/DBSJ
 A/Molecule type: mRNA
 A/Residues: 1-1081 <ING>
 A/Cross-references: EMBL:Y11795; NID:g2190349; PIDN:CAA72491.1; PID:g2190350
 A/Experimental source: ABA-treated callus
 C/Genetics:
 A/Gene: sps2
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: Glucosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |||
 Db 445 VIPPGMDFS 453

RESULT 10
 T04062
 sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
 C/Accession: T04062
 R/Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.;
 submitted to the Protein Sequence Database, March 1999
 A/Reference number: Z15184
 A/Accession: T04062
 A/Molecule type: DNA
 A/Residues: 1-1083 <BEV>
 A/Cross-references: EMBL:AL049487
 A/Experimental source: cultivar Columbia; BAC clone F28M11
 C/Genetics:
 A/Map position: 4
 A/Introns: 86/3; 116/3; 255/3; 322/2; 482/3; 526/3; 570/3; 629/3; 647/3; 668/3; 949/3; 5
 A/Note: F28M11.40
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 F:230-714/Domain: sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |||
 Db 483 VIPPGMDFS 491

RESULT 11
 T04103
 sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
 C/Species: Oryza sativa (rice)
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C/Accession: T04103
 R/Sakamoto, M.; Satozawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura, T.
 Plant Sci. 112, 207-217, 1995
 A/Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene that
 A/Reference number: Z15212
 A/Accession: T04103
 A/Status: preliminary; translated from GB/EMBL/DBSJ
 A/Molecule type: DNA
 A/Residues: 1-1084 <SAK>
 A/Cross-references: EMBL:D45890; PIDN:BA08304.1
 A/Experimental source: subsp. Japonica

C;Genetics:
 A;Gene: Spel
 A;Map position: 1
 A;Introns: 120/3; 200/2; 221/3; 452/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
 C;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C;Keywords: glycosyltransferase; hexosyltransferase
 F;196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 35; DB 2; Length 1084;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |::|::|::|
 Db 453 VVPXGMDYS 461

RESULT 12
 T24111
 hypothetical protein R10D12.10 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T24111
 R;Percy, C.
 Submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19842
 A;Accession: T24111
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-425 <WIL>
 A;Cross-references: EMBL:Z81109; PIDN: CAB03241.1; GSPDB: GN00023; CESP: R10D12.10
 C;Genetics:
 A;Gene: CESP:R10D12.10
 A;Map position: 5
 A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVFXGMDY 10
 |::|::|::|
 Db 335 EQVPGGLQY 344

RESULT 13
 H87660
 peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus
 C;Species: Caulobacter crescentus
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C;Accession: H87660
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of Caulobacter crescentus.
 A;Reference number: A87249; UID: 21173698; PMID: 11259647
 A;Accession: H87660
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-433 <STO>
 A;Cross-references: GB:AE005673; NID: g13425020; PIDN: AAK25284.1; GSPDB: GN00148
 C;Genetics:
 A;Gene: CC3322

Query Match 64.8%; Score 35; DB 2; Length 433;
 Best Local Similarity 54.5%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVFXGMDYS 11
 |::|::|::|
 Db 266 EVLPFGDYS 276

RESULT 14

H72784
 probable alkaline proteinase APE0263 - Aeropyrum pernix (strain K1)
 C;Species: Aeropyrum pernix
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C;Accession: H72784
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
 DNA Res. 6, 83-101, 1999
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
 A;Reference number: A72450; UID: 99310339; PMID: 10382966
 A;Accession: H72784
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-440 <KAW>
 A;Cross-references: DDBJ:AP000058; NID: G5103388; PIDN: BAA79178.1; FID: G5103657
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE0263
 C;Superfamily: subtilisin; subtilisin homology

Query Match 64.8%; Score 35; DB 2; Length 440;
 Best Local Similarity 66.7%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYVFXGMDY 10
 |::|::|::|
 Db 120 EVLPWGVY 128

RESULT 15
 T20173
 hypothetical protein C53A5.2 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C;Accession: T20173; T23857
 R;Mortimore, B.
 Submitted to the EMBL Data Library, November 1996
 A;Reference number: Z19232
 A;Accession: T20173
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1150 <WIL>
 A;Cross-references: EMBL:Z81486; PIDN: CAB03994.1; GSPDB: GN00023; CESP: C53A5.2
 A;Experimental source: clone C53A5
 R;Matthews, L.
 Submitted to the EMBL Data Library, August 1996
 A;Reference number: Z19808
 A;Accession: T23857
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1150 <WIL>
 A;Cross-references: EMBL:Z78015; PIDN: CAB01437.1; GSPDB: GN00023; CESP: C53A5.2
 A;Experimental source: clone R02D5
 C;Genetics:
 A;Gene: CESP: C53A5.2
 A;Map position: 5
 A;Introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2; 5

Query Match 64.8%; Score 35; DB 2; Length 1150;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |::|::|::|
 Db 562 VLPFGVDYS 570

Search completed: June 3, 2004, 12:00:00
 Job time: 10 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds

(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	70.4	1058	1	Q8RG86 fusobacteri
2	36	66.7	102	1	P31619 tobacco yel
3	36	66.7	1049	1	Q43802 oryza sativ
4	36	66.7	1068	1	P31927 zea mays (m
5	36	66.7	1081	1	Q04933 cratichneut
6	35	64.8	2778	1	P55824 drosophila
7	34.5	63.9	748	1	Q9nr64 homo sapien
8	34	63.0	154	1	P20423 oryza sativ
9	34	63.0	155	1	P08248 hordeum vul
10	34	63.0	168	1	P11970 populus nig
11	34	63.0	566	1	P94283 borrelia bu
12	34	63.0	908	1	Q8fj88 arabidopsis
13	34	63.0	908	1	Q8w4j9 arabidopsis
14	34	63.0	910	1	P59584 arabidopsis
15	33	61.1	276	1	Q58349 methanococ
16	33	61.1	283	1	Q5hvc9 pseudomonas
17	33	61.1	394	1	Q8kmv3 vibrio chol
18	33	61.1	421	1	P08503 rattus norv
19	33	61.1	421	1	Q82250 halomonas e
20	33	61.1	423	1	Q8zeu7 halomonas e
21	33	61.1	787	1	O60344 homo sapien
22	33	61.1	801	1	Q61851 mus muscullu
23	33	61.1	806	1	P18460 gallus gall
24	33	61.1	877	1	O74377 schizosacch
25	33	61.1	882	1	Q79399 simian foam
26	33	61.1	1401	1	P0C0344 vibrio chol
27	33	61.1	2717	1	P15822 homo sapien
28	32.5	60.2	472	1	P19102 xenopus lae
29	32	59.3	97	1	ET2A XENLA
30	32	59.3	150	1	PLAS DAUCA
31	32	59.3	165	1	Q88c90 pseudomonas
32	32	59.3	165	1	Q92bcs listeria in
33	32	59.3	175	1	TPX LISIN
					Q8y6u8 listeria mo
					Q04667 rattus norv
					HES3_RAT

34	32	59.3	231	1	ARAD ECOLI
35	32	59.3	231	1	ARAD_SALTY
36	32	59.3	233	1	HIS9_THEMA
37	32	59.3	288	1	CGD2_RAT
38	32	59.3	289	1	CGD2_HUMAN
39	32	59.3	289	1	CGD2_MOUSE
40	32	59.3	291	1	CGD1_BREARE
41	32	59.3	291	1	CGD1_XENLA
42	32	59.3	291	1	CGD2_CHICK
43	32	59.3	291	1	CGD2_XENLA
44	32	59.3	292	1	CGD2_CHICK
45	32	59.3	292	1	CGD3_HUMAN

ALIGNMENTS

RESULT 1

ID	CARB_FUSNN	STANDARD	PRT	1058 AA
AC	Q8RG86;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
GN	CARB OR FN0422.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;			
OC	Fusobacterium.			
OX	NCBI_TaxID=76856;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 25586;			
RX	MEDLINE=21886394; PubMed=11889109;			
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein N., Kyrpides N., Overbeek R.;			
RA	"Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";			
RL	J. Bacteriol. 184:2005-2018(2002).			
CC	- CAPALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.			
CC	- COPACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	- PATHWAY: Arginine biosynthesis.			
CC	- PATHWAY: Pyrimidine biosynthesis; first step.			
CC	- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).			
CC	- SIMILARITY: Belongs to the carb family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	EMBL; AB010554; AAL94625.1; ALT_INIT.			
DR	HAMAP; MF_01210; ; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; CPase_L.			
DR	InterPro; IPR005479; CPase_L_D2.			
DR	InterPro; IPR005480; CPase_L_D3.			
DR	InterPro; IPR005481; CPase_L_N.			
DR	InterPro; IPR004362; MGS-like.			
DR	Pfam; PF00289; CPasease_L_Chain; 2.			
DR	Pfam; PF02786; CPasease_L_D2; 2.			
DR	Pfam; PF02787; CPasease_L_D3; 1.			

```

DR Pfam; PF02142; MGS: 1.
DR PRINTS; PRO0098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAVOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT NP_BIND 352 382
FT METAL 284 288 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 3 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1B39F CRC64;

Query Match 70.4%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 9.4; 1; Indels 0; Gaps 0;
Matches 6; Conservative 3; Mismatches 1;

QY 2 EVVFXGMDYS 11
|:|:|:|:|
Db 190 EIVPENGINS 199

RESULT 2
Y1LK_TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN Vi.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
ON NCBI_TaxID=315599;
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81103; AAA47947.1; -.
CC FIRM; A42452; A42452.
CC InterPro; IPR002621; Gemini_mov.
CC Pfam; PF01708; Gemini_mov; I.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1B0AF55B67 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDYS 11
|:|:|:|:|
Db 190 EIVPENGINS 199

```

```

Db 7 QVVPFGINYS 16

RESULT 3
SPS_ORYSA STANDARD; PRT; 1049 AA.
ID SPS_ORYSA
AC Q43802;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14)
DE (UDP-glucose-fructose-phosphate glucosyltransferase).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Rhnartoidae; Oryzae; Oryza.
ON NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR36; TISSUE=Leaf;
RC MEDLINE=96235138; PubMed=8666248;
RA Valdez-Alarcon J.J., Ferrando M., Jimenez-Moraila B.,
RA Herrera-Estrella L.,
RT "Characterization of a rice sucrose-phosphate synthase-encoding
RT gene."
RL Gene 170:217-222(1996).
CC -!- FUNCTION: Involved in the regulation of carbon partitioning in the
CC leaves of plants. May regulate the synthesis of sucrose and
CC therefore play a major role as a limiting factor in the export of
CC photoassimilates out of the leaf.
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
CC sucrose 6-phosphate.
CC -!- ENZYME REGULATION: Activity regulated by phosphorylation and
CC moderated by concentration of metabolites and light.
CC -!- PATHWAY: Sucrose synthesis.
CC -!- SUBUNIT: Homodimer or homotetramer (By similarity).
CC -!- PTM: Phosphorylated. However, phosphorylation is not essential for
CC enzyme function (By similarity).
CC -!- SIMILARITY: Belongs to the glucosyltransferase family 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U31175; AAC49379.1; -.
CC FIRM; JCA783; JCA783.
CC Gramene; Q43802; -.
CC InterPro; IPR001296; Glyco_trans_1.
CC Pfam; PF00534; Glycosyltransferase; Phosphorylation.
KW Transferase; Glycosyltransferase; Phosphorylation.
FT DOMAIN 22 29 POLY-GLY.
FT DOMAIN 695 698 POLY-GLU.
FT DOMAIN 775 779 POLY-ARG.
SQ SEQUENCE 1049 AA; 116455 MW; ED862E2819AA4B04 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMDYS 11
|:|:|:|:|
Db 436 VVPPGMDYS 444

RESULT 4
SPS_MAIZE STANDARD; PRT; 1068 AA.
ID SPS_MAIZE
AC P31927;
DT 01-JUL-1993 (Rel. 26, Created)

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
phosphate glucosyltransferase).
DE SPS.
CN Zea mays (Maize).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
CC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 71-74; 206-212; 471-481 AND
RP 872-892.
RC STRAIN=CV, PIONEER 3184; TISSUE=Leaf;
RX MEDLINE=92338837; PubMed=1840396;
RA Worrell A.C., Bruneau J.-M., Summerfelt K., Boersig M., Voelker T.A.;
RT "Expression of a maize sucrose phosphate synthase in tomato alters
RT leaf carbohydrate partitioning."
RL Plant Cell 3:1121-1130(1991).
CC -!- FUNCTION: Involved in the regulation of carbon partitioning in the
CC leaves of plants. May regulate the synthesis of sucrose and
CC therefore play a major role as a limiting factor in the export of
CC photoassimilates out of the leaf.
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
CC sucrose 6-phosphate.
CC -!- ENZYME REGULATION: Activity regulated by phosphorylation and
CC moderated by concentration of metabolites and light.
CC -!- PATHWAY: Sucrose synthesis.
CC -!- SUBUNIT: Homodimer or homotetramer.
CC -!- DEVELOPMENTAL STAGE: Germinating seeds or mature leaves.
CC -!- PTM: Phosphorylated. However, phosphorylation is not essential for
CC enzyme function.
CC -!- SIMILARITY: Belongs to the glucosyltransferase family 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M57550; AA33513.1; -.
DR PIR; JQ1329; JQ1329.
DR MaizeDB; 25294; -.
DR InterPro; IPR001296; Glyco trans 1.
DR Pfam; PF00534; Glycos transf 1; 1.
DR Transferase; Glycosyltransferase; Phosphorylation.
KW DOMAIN 25 31 POLY-GLX.
FT DOMAIN 25 31
FT SEQUENCE 1068 AA; 118575 MW; 074679B5E9A1D282 CRC64;
SQ
Query Match 66.7%; Score 36; DB 1; Length 1068;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
Db 435 VIPPGMDFS 443
RESULT 5
ID SP82_CRAPL STANDARD; PRT; 1081 AA.
AC O4933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase 2 (EC 2.4.1.14) (UDP-glucose-fructose-
phosphate glucosyltransferase 2).
DE SP82.
GN Craterostigma plantagineum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamiids; Lamiales; Lamiales incertae sedis; Linderniaceae;
OC Craterostigma.
OX NCBI_TaxID=4153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97451773; PubMed=9306694;
RA Ingram J., Chandler J.W., Gallagher L., Salamini F., Bartels D.;
RT "Analysis of cDNA clones encoding sucrose-phosphate synthase in
RT relation to sugar interconversions associated with dehydration in the
RT resurrection plant *Craterostigma plantagineum* Hochst.";
RL Plant Physiol. 115:113-121(1997).
CC -!- FUNCTION: Involved in the regulation of carbon partitioning in the
CC leaves of plants. May regulate the synthesis of sucrose and
CC therefore play a major role as a limiting factor in the export of
CC photoassimilates out of the leaf.
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
CC sucrose 6-phosphate.
CC -!- ENZYME REGULATION: Activity regulated by phosphorylation and
CC moderated by concentration of metabolites and light.
CC -!- PATHWAY: Sucrose synthesis.
CC -!- SUBUNIT: Homodimer or homotetramer (By similarity).
CC -!- PTM: Phosphorylated. However, phosphorylation is not essential for
CC enzyme function (By similarity).
CC -!- SIMILARITY: Belongs to the glucosyltransferase family 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y11795; CAA72491.1; -.
DR PIR; T09837; T09837.
DR InterPro; IPR001296; Glyco trans 1.
DR Pfam; PF00534; Glycos transf 1; 1.
DR Transferase; Glycosyltransferase; Phosphorylation; Multigene family.
KW DOMAIN 245 248 POLY-SER.
FT DOMAIN 245 264
FT DOMAIN 256 264 POLY-GLU.
FT DOMAIN 787 790 POLY-ARG.
FT SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;
SQ
Query Match 66.7%; Score 36; DB 1; Length 1081;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
Db 445 VIPPGMDFS 453
RESULT 6
ID FAF_DROME STANDARD; PRT; 2778 AA.
AC P55824; Q9V9T6; Q9V0Z7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
DE (Ubiquitin thioesterase FAF) (Ubiquitin-specific processing protease
DE FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
DE FAF OR BCDNA:ID22582 OR CG1945.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
RC TISSUE=Eye imaginal disk;
RX MEDLINE=93202020; PubMed=1295747;
RX Fischer-Vize J.A., Rubin G.M., Lehmann R.;

SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2039AE6A CRC64;

Query Match 63.0%; Score 34; DB 1; Length 168;
 Best Local Similarity 54.5%; Pred. No. 9.5;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGMDYS 11
 : : : : :
 DB 112 EDVPSGMDVS 122

RESULT 11

SYTB_BORBU STANDARD; PRT; 566 AA.

ID SYTB_BORBU STANDARD; PRT; 566 AA.

AC E94283;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--

DE tRNA ligase beta chain) (PheRS).

DE PHE1 OR B50514.

OS Borrelia burgdorferi (Lyme disease spirochete).

GN Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.

OC NCBI_TaxID=139;

OX NCBI_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RA Barbour A.G., Hinnebusch J.;

RT "Phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and

RT thioridoxin reductase gene of *Borrelia burgdorferi*."

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,

RA Peterson J., Kervatage A.R., Quackenbush J., Salzberg S., Hanson M.,

RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,

RA Uterback T., Watthey L., McDonald L., Artach P., Bowman C.,

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,

RA Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochete, *Borrelia*

RT *burgdorferi*."

RL Nature 390:580-586 (1997).

CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +

CC diphosphate + L-phenylalanyl-tRNA(Phe).

CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By

CC similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain

CC family. Subfamily 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC EMBL; U82978; BAB41019.1; -

CC EMBL; AE001153; AAC66870.1; -

CC PIR; A70164; A70164.

CC TIGR; BB0514; -

CC HAMAP; MF 00284; -; 1.

CC InterPro; IPR005147; B5.

CC InterPro; IPR004531; PheT_arch.

CC Pfam; PF03484; B5; 1.

CC TIGRFAMs; TIGR00471; pHeT_arch; 1.

CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.

KW


```

FT DOMAIN 146 459 NB-ARC.
FT REPEAT 575 599 LRR 1.
FT REPEAT 600 623 LRR 2.
FT REPEAT 842 867 LRR 3.
FT NP_BIND 192 199 ATP (POTENTIAL).
SQ SEQUENCE 908 AA; 104448 MW; 3111991B17239693 CRC64;

Query Match 63.08; Score 34; DB 1; Length 908;
Best Local Similarity 60.08; Pred. NO. 54;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVFXGMDY 10
Db 883 EKLPGGEDY 892
[::|||]

RESULT 13
RPP8_ARATH STANDARD; PRT; 908 AA.
AC Q8W4J9; Q8GMG5; Q9MSA1; Q9ZSY3; Q9ZSY4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Disease resistance protein RPP8 (Resistance to Peronospora parasitica
DE protein 8).
GN RPP8 OR HRT OR AT5G43470 OR MWF20.19.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND
RP RPP8-3, AND VARIANTS.
RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
RX MEDLINE=99030193; PubMed=9811794;
RA McDowell J.B., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intragenic recombination and diversifying selection contribute to the
RT evolution of downy mildew resistance at the RPP8 locus of
RT Arabidopsis."
RL Plant Cell 10:1861-1874 (1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
RC STRAIN=cv. Di-17;
RX MEDLINE=20271766; PubMed=10810142;
RA Cooley M.B., Pathirana S., Wu H.-J., Kachroo P., Klessig D.F.;
RT "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer
RT resistance to both viral and oomycete pathogens."
RL Plant Cell 12:663-676 (2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and PAC
RT clones."
RL DNA Res. 7:31-63 (2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.F.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Chook R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayaishizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,

```

```

RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tame R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome."
RL Science 302:842-846 (2003).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayaishizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [6]
RP INTERACTION WITH TIP.
RX MEDLINE=20496823; PubMed=11041886;
RA Ren T., Qu F., Morris T.J.;
RT "HRT gene function requires interaction between a NAC protein and
RT viral capsid protein to confer resistance to turnip crinkle virus."
RL Plant Cell 12:1917-1926 (2000).
CC -!- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. The interaction
CC with TIP (TCV-interacting protein) may be essential for the
CC recognition of the avirulence proteins, and the triggering of the
CC defense response.
CC -!- SUBUNIT: Interacts with the NAC protein TIP.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8W4J9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8W4J9-2; Sequence=VSP 007171, VSP 007172;
CC Note=Has been shown to exist only in cv. Columbia so far;
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition.
CC -!- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
CC cv. Columbia are probably due to an unequal crossing-over between
CC the highly related RPP8 and RPP8A genes present in cv. Landsberg
CC erecta. Such variations probably modify the specificity of
CC pathogen recognition.
CC -!- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
CC specifically recognizes the Emco5 avirulence protein from
CC Peronospora parasitica, while it is not the case in cv. Di-17,
CC where it confers resistance to Turnip Crinkle Virus upon
CC recognition of the viral capsid protein.
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME=NTB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://nblrrs.ucdavis.edu".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF089710; AAC83165.1; -
CC EMBL; AF089711; AAC78631.1; -
CC EMBL; AF234174; BAA36987.1; -
CC EMBL; AB025638; BAA97426.1; -
CC EMBL; AY062514; AAL32592.1; -
CC EMBL; AK118862; BAC43449.1; -
CC InterPro; IPR000767; Disease_resist.

```



```

RESULT 15
Y939 METJA          STANDARD;          PRT;    276 AA.
AC 058349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0939.
GN MJ0939.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67537; AAB98946.1; -.
DR PIR; C64417; C64417.
DR TIGR; MJ0939; -.
DR InterPro; IPR001163; sRNP_Sm.
DR InterPro; IPR005358; UPF0153.
DR Pfam; PF03692; UPF0153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 33454 MW; 97BD69D392BC8FDF CRC64;

Query Match          61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred. No. 26;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
Db      141 EBIENGMEHS 151

Search completed: June 3, 2004, 11:49:52
Job time : 5.86667 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EEWVPXGMDYS 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	3	Q12479 saccharomy
2	38	70.4	319	16	Q85V7 oceanobacil
3	38	70.4	363	17	Q30260 archaeoglob
4	38	70.4	1044	16	Q8DIH0 synechococc
5	36	66.7	149	16	Q82ZB4 enterococc
6	36	66.7	341	10	Q22081 citrus unsh
7	36	66.7	348	10	Q22096 citrus unsh
8	36	66.7	452	10	Q8W568 arabidopsis
9	36	66.7	460	10	Q9C9T7 arabidopsis
10	36	66.7	1047	10	P93782 saccharum o
11	36	66.7	1083	10	Q9SN30 arabidopsis
12	36	66.7	1084	10	Q43010 oryza sativ
13	36	66.7	1100	10	Q8S064 oryza sativ
14	35	64.8	219	5	Q9GQ04 eriocheir s
15	35	64.8	253	16	Q8XPAB clostridium
16	35	64.8	298	2	O52367 rhizobium t

17	35	64.8	308	16	Q8CCO0 pseudomonas
18	35	64.8	368	2	Q8KL43 rhizobium e
19	35	64.8	425	5	Q9XVK4 caenorhabdi
20	35	64.8	433	16	Q9A382 caulobacter
21	35	64.8	440	17	Q9YFI3 aeropyrum p
22	35	64.8	474	16	Q83192 shigella fl
23	35	64.8	511	2	O52680 escherichia
24	35	64.8	517	16	Q8XZL5 ralstonia s
25	35	64.8	595	10	Q94617 andirographi
26	35	64.8	745	5	Q95P46 carcinus ma
27	35	64.8	1031	5	Q9U6A3 callinectes
28	35	64.8	1150	5	O17704 caenorhabdi
29	35	64.8	1410	2	O52673 escherichia
30	35	64.8	1420	2	O52666 escherichia
31	35	64.8	1474	17	Q27146 methanobact
32	35	64.8	1828	16	Q98K29 rhizobium l
33	34.5	63.9	748	4	Q8TBJ7 homo sapien
34	34	63.0	215	16	Q8R9L5 thermoanaer
35	34	63.0	222	16	Q8ESL9 streptococc
36	34	63.0	222	16	Q8DZW9 streptococc
37	34	63.0	245	16	Q7V6Q4 prochlorococ
38	34	63.0	251	9	Q855K6 mycobacteri
39	34	63.0	284	2	Q8GP33 lactobacilli
40	34	63.0	290	16	Q8U7J0 agrobacteri
41	34	63.0	296	17	Q9YET8 aeropyrum p
42	34	63.0	337	16	Q7U9F5 rhodospirell
43	34	63.0	344	16	Q815A7 bacillus ce
44	34	63.0	357	17	O29920 archaeoglob
45	34	63.0	366	17	O29451 archaeoglob

ALIGNMENTS

RESULT 1

Q12479	PRELIMINARY;	PRT;	156 AA.
ID	Q12479	PRELIMINARY;	PRT;
AC	Q12479;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DR	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)	
DE	ORF YOR013W.		
GN	YOR013W.		
OS	Saccharomyces cerevisiae (Baker's Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	De haan M., Grivell L.A., Maarse A.C.;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	MIPS;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FY1679;		
RA	De haan M., Maarse A.C., Grivell L.A.;		
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=FY1679;		
RX	MEDLINE=94019318; PubMed=8413243;		
RA	Dumont M.E., Schlichter J.B., Cardillo T.S., Hayes M.K., Bethlenny G.,		
RA	Sherman F.;		
RT	"CIC2 encodes a factor involved in mitochondrial import of yeast		
RT	cytochrome C.;"		
RL	Mol. Cell. Biol. 13:6442-6451(1993).		
RN	[5]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=FY1679;		
RC	STRAIN=FY1679;		
RX	MEDLINE=94169519; PubMed=7764548;		

RA Lee Y.S., Shimizu J., Yoda K., Yamaaki M.;
 RT "Molecular cloning of a gene, DHS1, which complements a drug-
 RL hypersensitive mutation of the yeast *Saccharomyces cerevisiae*.";
 RA Blotcl. Biotechnol. Biochem. 58:391-395 (1994).
 DR EMBL; 274920; CAA99201.1; -;
 DR EMBL; X87331; CAA60762.1; -;
 DR PIR; S54619; S54619;
 DR SGD; S0005539; YOR013W.
 SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;

Query Match 74.1%; Score 40; DB 3; Length 156;
 Best Local Similarity 77.8%; Pred. No. 3;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMDY 10
 |||:||||
 DB 50 EWMPLGMDY 58

RESULT 2

Q8ESV7 PRELIMINARY; PRT; 319 AA.
 AC Q8ESV7;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Hypothetical conserved protein.
 GN O80509.
 OS Oceanobacillus iheyensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;
 RX MEDLINE=22220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments";
 RL Nucleic Acids Res. 30:3927-3935 (2002).
 DR EMBL; AF004594; BAC12465.1; -;
 DR InterPro; IPR001279; Blactamase-like.
 DR Pfam; PF00753; lactamase B; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 319 AA; 35617 MW; 3BDAE4BF13E79E37 CRC64;

Query Match 70.4%; Score 38; DB 16; Length 319;
 Best Local Similarity 60.0%; Pred. No. 18;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
 |||:||||
 DB 189 EQLVPHGIDY 198

RESULT 3

O30260 PRELIMINARY; PRT; 363 AA.
 AC O30260;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Hypothetical protein Af2411.
 GN AF2411.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370 (1997).
 DR EMBL; AE001109; AAS91255.1; -;
 DR PIR; D69551; D69551.
 DR TIGR; AF2411; -;
 DR InterPro; IPR002103; Bac_luciferase.
 DR Pfam; PF00296; bac_luciferase; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;

Query Match 70.4%; Score 38; DB 17; Length 363;
 Best Local Similarity 54.5%; Pred. No. 20;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||:||||
 DB 120 ENIVPYGIDFS 130

RESULT 4

Q8DIH0 PRELIMINARY; PRT; 1044 AA.
 AC Q8DIH0;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLL1618.
 OS *Synechococcus elongatus* (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT *Thermosynechococcus elongatus* BP-1";
 RL DNA Res. 9:123-130 (2002).
 DR EMBL; AP005374; BAC09170.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001036; Acrflavin_res.
 DR InterPro; IPR004764; HAE1.
 DR Pfam; PF00873; Acr_tran; i.
 DR PRINTS; PRO0702; ACRIFLAVINRP.
 DR TIGRFAMs; TIGR00915; 2A0602; 1.
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 70.4%; Score 38; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 67;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||:||||
 DB 843 EEVLPNGIGYS 853

```

RESULT 5
Q82ZB4 PRELIMINARY; PRT; 149 AA.
AC Q82ZB4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Large conductance mechanosensitive channel protein.
GN MSLC OR EF3152.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Dougherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016956; AAO82828.1; -.
DR TIGR; EF3152; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005216; F:ion channel activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR001185; MS_channel.
DR Pfam; PF01741; MscL; 1.
DR PRINTS; PR01264; MECHCHANNEL.
DR ProDom; PD007253; MS_channel; 1.
DR TIGRFAMs; TIGR00220; mscL; 1.
KW Complete proteome.
SQ SEQUENCE 149 AA; 16127 MW; 555799BF1E47D34E CRC64;

Query Match 66.7%; Score 36; DB 16; Length 149;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGMDY 10
DB 113 EVVVPXGMDY 122

RESULT 6
Q22081 PRELIMINARY; PRT; 341 AA.
AC Q22081;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
GN CITSPS2.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RX Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; ABC06660; BAA22071.1; -.
DR PIR; S72650; S72650.
FT NON_TER 1 341
FT NON_TER 348
SQ SEQUENCE 348 AA; 38556 MW; BE1C21EBA6FF5C5E CRC64;

Query Match 66.7%; Score 36; DB 10; Length 348;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 234 VVPXGMDYS 242

RESULT 8
Q8W568 PRELIMINARY; PRT; 452 AA.
AC Q8W568;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE At1g73750/F25P22_17.

```

```

RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RX Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; ABC06660; BAA22071.1; -.
DR PIR; S72649; S72649.
FT NON_TER 1 341
FT NON_TER 341
SQ SEQUENCE 341 AA; 38136 MW; 61417A69C4560777 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 341;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 228 VVPXGMDYS 236

RESULT 7
Q22096 PRELIMINARY; PRT; 348 AA.
AC Q22096;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
GN CITSPS3.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Gen. Genet. 252:346-351(1996).
RN [2];
RP SEQUENCE FROM N.A.
RC TISSUE=Juice sacs and segment epidermis;
RX Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; ABC06660; BAA22071.1; -.
DR PIR; S72650; S72650.
FT NON_TER 1 348
FT NON_TER 348
SQ SEQUENCE 348 AA; 38556 MW; BE1C21EBA6FF5C5E CRC64;

Query Match 66.7%; Score 36; DB 10; Length 348;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 234 VVPXGMDYS 242

RESULT 8
Q8W568 PRELIMINARY; PRT; 452 AA.
AC Q8W568;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE At1g73750/F25P22_17.

```

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RA "Arabidopsis cDNA clones";
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF419606; RAL31938.1; --
 DR ENBL; AY113044; AM47352.1;
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR GO; GO:0004287; F: prolyl oligopeptidase activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estra.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 SQ SEQUENCE 452 AA; 49682 MW; A159955B21742C4A CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 452;
 Best Local Similarity 70.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDY 10
 ||| ||| |||
 Db 210 EEDVPSANDY 219
 RESULT 9
 Q9C9T7 PRELIMINARY; PRT; 460 AA.
 AC Q9C9T7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DB Hypothetical protein.
 GN F25P22.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=cv. Columbia;
 RC MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Creasy B., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Lancin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzgier S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RT Nature 408:816-820 (2000).
 RL ENBL; AC012679; AGS2073.1; --
 DR PIR; G96764; G96764.
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR GO; GO:0004287; F: prolyl oligopeptidase activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estra.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 460 AA; 50564 MW; E94B27B5C4B249EC CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 460;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDY 10
 ||| ||| |||
 Db 218 EEDVPSANDY 227
 RESULT 10
 P93782 PRELIMINARY; PRT; 1047 AA.
 ID P93782;
 AC P93782;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DB Sucrose-phosphate synthase (EC 2.4.1.14) (Fragment).
 GN SCSP81.
 OS Saccharum officinarum (Sugarcane).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 OX NCBI_TaxID=4547;
 [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Sugiharto B., Sakakibara H., Sugiyama T.;
 RA "Differential Expression of Two Genes for Sucrose-Phosphate Synthase
 in Sugarcane: Molecular Cloning of the cDNAs and Comparative Analysis
 of Gene Expression";
 RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RL ENBL; AB001337; BAA19241.1; --
 DR GO; GO:0046524; F: sucrose-phosphate synthase activity; IEA.
 DR GO; GO:0016757; F: transferase activity; IEA.
 DR GO; GO:0009058; P: biosynthesis; IEA.
 DR InterPro; IPR001296; Glyco_transf.1.
 DR Pfam; PF00534; Glycosyltransferase; 1.
 KW Glycosyltransferase; Transferase.
 FT NON_TER 1
 RN TER 1
 SQ SEQUENCE 1047 AA; 116379 MW; D0EBDB34961E1D83D CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 1047;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11
 ||| ||| |||
 Db 414 VVPXGMDYS 422
 RESULT 11

Q9SN30
ID Q9SN30 PRELIMINARY; PRT; 1083 AA.
AC Q9SN30
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
GN P28M11.40 OR A74G10120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.,
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049487; CAB39764.1; -.
DR EMBL; AL161516; CAB78135.1; -.
DR PIR; T04062; T04062.
DR GO; GO:0045524; F:sucrose-phosphate synthase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009058; F:biosynthesis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1083 AA; 122688 MW; EFC126FCA2137BB1 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 483 VIPPGWDFS 491

RESULT 12
Q43010
ID Q43010 PRELIMINARY; PRT; 1084 AA.
AC Q43010
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Sucrose phosphate synthase.
GN SPS1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Japonica;
RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shimada H.,
RA Fujimura T.,
RT "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS)
RT gene that is specifically expressed in the source organ."
RL Plant Sci. 112:207-217(1995).
DR EMBL; D45890; BAA08304.1; -.
DR PIR; T04103; T04103.
DR Gramene; Q43010; -.
DR GO; GO:0044559; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0009058; F:biosynthesis; IEA.

DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycos_transf_1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1084 AA; 119417 MW; 4D079AFEDD2F92B CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 453 VIPPGWDFS 461

RESULT 13
Q8S064
ID Q8S064 PRELIMINARY; PRT; 1100 AA.
AC Q8S064
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Putative sucrose-phosphate synthase.
GN P0678F11.14.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.,
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0678F11."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003437; BAB86107.1; -.
DR Gramene; Q8S064; -.
DR GO; GO:0044559; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0009058; F:biosynthesis; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycos_transf_1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1100 AA; 121170 MW; 084F4604BA389CAD CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1100;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 469 VIPPGWDFS 477

RESULT 14
Q9GQ04
ID Q9GQ04 PRELIMINARY; PRT; 219 AA.
AC Q9GQ04
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Na+/K+/2Cl-cotransporter (Fragment).
OS Eriocheir sinensis (Chinese mitten crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Subrachyura; Eucarida; Varunidae; Eriocheir.
OX NCBI_TaxID=95602;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gill;

RA Weihrach D., Towle D.W.;
RT "Na+/H+-exchanger and Na+/K+/2Cl- -cotransporter are expressed in
RL gills of the euryhaline Chinese crab *Eriocheir sinensis*.";
RT Comp. Biochem. Physiol. 126:S158-S158(2000).
DR EMBL; AF301160; AAG39938.1; -.
FT NON_TER 1
FT NON_TER 219 219
SQ SEQUENCE 219 AA; 24159 MW; 599442DA26F0D3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VEPXGMDYS 11
DB 107 VPQGLDYS 114

RESULT 15

Q8XPA8 PRELIMINARY; PRT; 253 AA.
AC Q8XPA8;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein CPE0057.
GN CPE0057.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
CX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
EX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79763.1; -.
DR GO; GO:0003824; F.catalytic activity; IEA.
DR InterPro; IPR009036; MoEB.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000594; ThiF_domain.
DR Pfam; PF00899; ThiF; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 27714 MW; 814DF79D3E0D7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
DB 108 EEIIPDDVDY 117

Search completed: June 3, 2004, 11:57:32
Job time : 30.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	5 ABB80525	Abb80525 Hepatitis
2	50	96.2	11	5 ABB80521	Abb80521 Hepatitis
3	50	96.2	11	5 ABB80522	Abb80522 Hepatitis
4	50	96.2	11	5 ABB80566	Abb80566 Hepatitis
5	50	96.2	11	5 ABB80563	Abb80563 Hepatitis
6	50	96.2	11	5 ABB80565	Abb80565 Hepatitis
7	50	96.2	11	5 ABB80567	Abb80567 Hepatitis
8	50	96.2	11	5 ABB80559	Abb80559 Hepatitis
9	50	96.2	11	5 ABB80526	Abb80526 Hepatitis
10	50	96.2	11	5 ABB80564	Abb80564 Hepatitis
11	50	96.2	11	5 ABB80568	Abb80568 Hepatitis
12	46	88.5	11	5 ABB80561	Abb80561 Hepatitis
13	46	88.5	11	5 ABB80524	Abb80524 Hepatitis
14	46	88.5	11	5 ABB80529	Abb80529 Hepatitis
15	46	88.5	11	5 ABB80528	Abb80528 Hepatitis
16	46	88.5	11	5 ABB80562	Abb80562 Hepatitis
17	45	86.5	11	5 ABB80523	Abb80523 Hepatitis
18	45	86.5	11	5 ABB80536	Abb80536 Hepatitis
19	45	86.5	11	5 ABB80558	Abb80558 Hepatitis
20	45	86.5	11	5 ABB80560	Abb80560 Hepatitis
21	45	86.5	11	5 ABB80527	Abb80527 Hepatitis
22	45	86.5	11	5 ABB80535	Abb80535 Hepatitis
23	45	86.5	11	5 ABB80540	Abb80540 Hepatitis
24	45	86.5	11	5 ABB80539	Abb80539 Hepatitis
25	44	84.6	11	5 ABB80549	Abb80549 Hepatitis

26	44	84.6	11	5 ABB80544	Abb80544 Hepatitis
27	44	84.6	11	5 ABB80553	Abb80553 Hepatitis
28	44	84.6	11	5 ABB80552	Abb80552 Hepatitis
29	44	84.6	11	5 ABB80545	Abb80545 Hepatitis
30	42	80.8	11	5 ABB80530	Abb80530 Hepatitis
31	41	78.8	11	5 ABB80542	Abb80542 Hepatitis
32	41	78.8	11	5 ABB80543	Abb80543 Hepatitis
33	41	78.8	11	5 ABB80538	Abb80538 Hepatitis
34	40	76.9	11	5 ABB80548	Abb80548 Hepatitis
35	40	76.9	11	5 ABB80547	Abb80547 Hepatitis
36	40	76.9	11	5 ABB80556	Abb80556 Hepatitis
37	40	76.9	11	5 ABB80557	Abb80557 Hepatitis
38	40	76.9	11	5 ABB80537	Abb80537 Hepatitis
39	40	76.9	11	5 ABB80551	Abb80551 Hepatitis
40	40	76.9	11	5 ABB80541	Abb80541 Hepatitis
41	40	76.9	20	2 AAU76810	Aau76810 Hepatitis
42	40	76.9	1022	4 ABG03621	Abg03621 Novel hum
43	40	76.9	1022	4 ABG08173	Abg08173 Novel hum
44	40	76.9	1022	4 ABG05826	Abg05826 Novel hum
45	39	75.0	11	5 ABB80546	Abb80546 Hepatitis

ALIGNMENTS

RESULT 1

ABB80525
ID ABB80525 standard; peptide; 11 AA.

AC ABB80525;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WO-US023169.

21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.

Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 2
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.
 XX
 AC ABB80521;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 SQ The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 2
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.
 XX
 AC ABB80521;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 SQ The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 2
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.
 XX
 AC ABB80521;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 SQ The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 2
 ABB80522
 ID ABB80522 standard; peptide; 11 AA.
 XX
 AC ABB80522;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 SQ The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 2
 ABB80522
 ID ABB80522 standard; peptide; 11 AA.
 XX
 AC ABB80522;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 SQ The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

ABB80566
ID ABB80566 standard; peptide; 11 AA.
XX AC ABB80566;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Valyl carbonyl forming keto-amide linkage with
FT Modified-site 11 residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
XX Query Match 96.2%; Score 50; DB 5; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.002;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB |||||
1 BEVVPXGMSYS 11
RESULT 6
ABB80565
ID ABB80565 standard; peptide; 11 AA.
XX AC ABB80565;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norleucyl carbonyl forming keto-amide linkage
FT with residue 7"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
XX Query Match 96.2%; Score 50; DB 5; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.002;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB |||||
1 BEVVPXGMSYS 11
RESULT 5
ABB80563
ID ABB80563 standard; peptide; 11 AA.
XX AC ABB80563;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX	(CORV-) CORVAS INT INC.
PA	Lim-Wilby M, Levy OE, Brunck TK;
XX	WPI; 2002-361643/39.
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
PT	Claim 17; Page 65; 69pp; English.
DR	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
CC	Sequence 11 AA;
PS	Query Match 96.2%; Score 50; DB 5; Length 11;
BS	Best Local Similarity 100.0%; Pred. No. 0.002;
XX	Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 EEVVPXGMSYS 11
DB	1 EEVFPXGMSYS 11
RESULT 8	
ABB80559	ID ABB80559 standard; peptide; 11 AA.
XX	AC ABB80559;
DT	08-OCT-2002 (first entry)
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
KW	Synthetic.
OS	
XX	Key Location/Qualifiers
FH	1 /note= "N-terminal acetyl"
FT	6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	8 /note= "Oxymethionine"
FT	8 /note= "D-form residue"
FT	11 /note= "C-terminal amide"
FT	
PN	WO200208251-A2.
XX	31-JAN-2002.
XX	19-JUL-2001; 2001WO-US023169.
PF	21-JUL-2000; 2000US-0220101P.
PR	(CORV-) CORVAS INT INC.
XX	Lim-Wilby M, Levy OE, Brunck TK;
PA	WPI; 2002-361643/39.
XX	
DR	

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
|||||
RESULT 9
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX
AC ABB80526;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT WO200208251-A2.
XX
PN 31-JAN-2002.
XX
PD 19-JUL-2001; 2001WO-US023169.
XX
PF 21-JUL-2000; 2000US-0220101P.
XX
PR (CORV-) CORVAS INT INC.
XX
PA Lim-Wilby M, Levy OE, Brunck TK;
XX
PI WPI; 2002-361643/39.
XX
DR Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
PS Claim 17; Page 64; 69pp; English.
XX
SQ

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
|||||
RESULT 10
ABB80564
ID ABB80564 standard; peptide; 11 AA.
XX
AC ABB80564;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
XX
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT WO200208251-A2.
XX
PN 31-JAN-2002.
XX
PD 19-JUL-2001; 2001WO-US023169.
XX
PF 21-JUL-2000; 2000US-0220101P.
XX
PR (CORV-) CORVAS INT INC.
XX
PA Lim-Wilby M, Levy OE, Brunck TK;
XX
PI WPI; 2002-361643/39.
XX
DR Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
PS Claim 17; Page 65; 69pp; English.
XX
SQ The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.

AC ABB80568;
XX
DT 08-OCT-2002 (first entry)

XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
DE	
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
KW	
XX	Synthetic.
OS	

RESULT 12	
ABB80561	
ID ABB80561	standard; peptide; 11 AA.
XX	
XX	
AC	
ABB80561;	
XX	
XX	
DT	
XX	
08-OCT-2002	(first entry)
DE	
Hepatitis C virus NS3/NS4a	serine protease inhibitor peptide #41.
XX	
Hepatitis C virus; HCV;	serine protease; inhibitor; alpha-ketoamide;
KW	virucide.
XX	
XX	
os	Synthetic.

XX	Key	Location/Qualifiers
FH	Modified-site	1
FT		

FT	Modified-site	6	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT		8	
FT	Modified-site		/note= "Oxymethionine"
FT		8	
FT	Misc-difference		/note= "D-form residue"
FT		11	
FT	Modified-site		/note= "C-terminal amide"
FT			

RESULT 13
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
XX ABB80524;
XX

	Matches	11; Conservativ
Qy	1	EEVVPXGMSYS 11
pb	1	EEVVPXGMSYS 11

Matches	11;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
Qy	1	EVVVPXGMSYS	11					
Dd	1	EVVVPXGMSYS	11					

DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
 DE
 XX
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX
 XX WPI; 2002-361643/39.
 DR
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 FT
 XX
 XX Claim 17; Page 64; 69pp; English.
 PS
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 XX Sequence 11 AA;
 SQ
 Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0;
 QY 1 EEVVPXGMSYS 11
 DB 1 EEVVPXGMDYS 11
 RESULT 14
 ABB80529
 ID ABB80529 standard; peptide; 11 AA.
 XX
 AC ABB80529;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
 DE
 XX
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX
 XX WPI; 2002-361643/39.
 DR
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 FT
 XX
 XX Claim 17; Page 64; 69pp; English.
 PS
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 XX Sequence 11 AA;
 SQ
 Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0;
 QY 1 EEVVPXGMSYS 11
 DB 1 EEVVPXGMDYS 11
 RESULT 15
 ABB80528
 ID ABB80528 standard; peptide; 11 AA.
 XX
 AC ABB80528;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
 DE
 XX
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"

FT Misc-difference 8 residue 7"
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 XX 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 XX
 XX (CORV-) CORVAS INT INC.
 XX
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 64; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMSYS 11
 | | | | | | | | | |
 Db 1 EEVVPXGMDYS 11
 | | | | | | | | | |
 Search completed: June 3, 2004, 11:48:23
 Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-9
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A_COMB.psp:
2: /cgn2_6/ptodata/2/iaa/5B_COMB.psp:
3: /cgn2_6/ptodata/2/iaa/6A_COMB.psp:
4: /cgn2_6/ptodata/2/iaa/6B_COMB.psp:
5: /cgn2_6/ptodata/2/iaa/FCRUS_COMB.psp:
6: /cgn2_6/ptodata/2/iaa/backfiles1.psp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	3472	4 US-09-408-020-4	Sequence 4, Appli
2	36	69.2	382	4 US-09-134-000C-3738	Sequence 3738, Ap
3	34	65.4	947	4 US-09-228-988-73	Sequence 73, Appl
4	34	65.4	1191	4 US-09-540-236-2902	Sequence 2902, Ap
5	34	65.4	1407	4 US-09-328-352-7885	Sequence 7885, Ap
6	33	63.5	12	4 US-09-760-946-2	Sequence 2, Appli
7	33	63.5	12	4 US-09-760-946-3	Sequence 3, Appli
8	33	63.5	45	2 US-08-637-7598-236	Sequence 236, App
9	33	63.5	45	3 US-08-871-355A-236	Sequence 236, App
10	33	63.5	45	4 US-09-201-945-236	Sequence 236, App
11	33	63.5	65	6 5177197-51	Patent No. 5177197
12	33	63.5	410	6 5177197-1	Patent No. 5177197
13	33	63.5	1394	6 5177197-30	Patent No. 5177197
14	32	61.5	10	3 US-09-357-952-66	Sequence 66, Appl
15	32	61.5	10	4 US-09-521-650-66	Sequence 66, Appl
16	32	61.5	102	2 US-08-168-888-66	Sequence 66, Appl
17	32	61.5	152	2 US-08-580-988A-23	Sequence 23, Appl
18	32	61.5	152	3 US-08-460-694-4	Sequence 4, Appli
19	32	61.5	152	3 US-08-460-744-4	Sequence 4, Appli
20	32	61.5	152	3 US-07-667-711B-4	Sequence 7, Appli
21	32	61.5	173	1 US-08-193-977-7	Sequence 7, Appli
22	32	61.5	189	2 US-08-464-517-21	Sequence 21, Appl
23	32	61.5	189	2 US-08-246-361A-21	Sequence 21, Appl
24	32	61.5	189	3 US-08-463-772-21	Sequence 21, Appl
25	32	61.5	189	5 PCT-US93-05000-21	Sequence 21, Appl
26	32	61.5	236	2 US-08-464-517-22	Sequence 22, Appl
27	32	61.5	236	2 US-08-246-361A-22	Sequence 22, Appl

Sequence 22, Appl
Sequence 22, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 23, Appl
Sequence 6, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 2, Appl
Sequence 19, Appl
Sequence 20, Appl
Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632937

GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/09/408.020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 73.1%; Score 38; DB 4; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
Db 2294 EDVIRGIGSFS 2304

RESULT 2
US-09-134-000C-3738
; Sequence 3738, Application US/09134000C
; Patent No. 6617156
GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3738
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Enterococcus faecalis

```
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: {327}..(328)
; OTHER INFORMATION: Amino acids 327 & 328 are Xaa wherein Xaa = any amino acid.
US-09-134-000C-3738

Query Match      69.2%; Score 36; DB 4; Length 382;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 332 LIPEGMSYS 340

RESULT 3
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228.986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match      65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 686 VMPGSGISYS 694

RESULT 4
US-09-540-236-2902
; Sequence 2902, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2902
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2902

Query Match      65.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 783 EILPVGWAY 791

RESULT 5
US-09-328-352-7885
```

```
; Sequence 7885, Application US/09328352
; Patent No. 8562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match      65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 596 EVPEGLSP 604

RESULT 6
US-09-760-946-2
; Sequence 2, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghro, Elise
; APPLICANT: Goudreau, Nathalie
; APPLICANT: Halmos, Teddy
; APPLICANT: Llinas-Brunet, Montse
; TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
; FILE REFERENCE: 13/076-1-C1
; CURRENT APPLICATION NUMBER: US/09/760.946
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/542,675
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/128,011
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Substrate for recombinant HCV NS3 protease radiometric assay
US-09-760-946-2

Query Match      63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
Db 1 DDIVPCNSYT 11

RESULT 7
US-09-760-946-3
; Sequence 3, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghro, Elise
```

APPLICANT: Goudreau, Nathalie
APPLICANT: Halmos, Teddy
APPLICANT: Llinas-Brunet, Montse
TITLE OF INVENTION: Macrocytic Peptides Active Against the Hepatitis C Virus
FILE REFERENCE: 13/076-1-C1
CURRENT APPLICATION NUMBER: US/09/760,946
CURRENT FILING DATE: 2001-08-23
PRIOR APPLICATION NUMBER: US 09/542,675
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: US 60/128,011
PRIOR FILING DATE: 1999-04-06
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 12
TYPE: PRI
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tracer for NS3 protease assay
NAME/KEY: MOD RES
LOCATION: (1)
OTHER INFORMATION: Asp at position 1 is biotinylated
NAME/KEY: MOD RES
LOCATION: (10)
OTHER INFORMATION: Tyr at position 10 is iodinated with I-125
US-09-760-946-3

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
Db 1 DDIVPCMSYT 11

RESULT 8
US-08-637-759B-236
Sequence 236, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:

SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 1 BEISPLGWSY 10

RESULT 9
US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 1 BEISPLGWSY 10

RESULT 10
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holder
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPLS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
: : : : :
DB 1 EEISPLGWSY 10

RESULT 11
5177197-51
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSENS-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 51
; LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 12;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
: : : : :
DB 52 KEICPGGNGYT 62

RESULT 12
5177197-1
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSENS-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 1
; LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 97;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
: : : : :
DB 399 KEICPGGNGYT 409

RESULT 13
5177197-30
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSENS-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 30
; LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
: : : : :
DB 399 KEICPGGNGYT 409

RESULT 14
US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whole
; TITLE OF INVENTION: Fluorescence screening Assays for Caspases, Peptidases, Proteases
; TITLE OF INVENTION: Other Enzymes and the Use Thereof
; FILE REFERENCE: 1735.0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642

; EARLIER FILING DATE: 21-JUL-1998
 ; NUMBER OF SEQ ID NOS: 139
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 66
 ; LENGTH: 10
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
 ; OTHER INFORMATION: Peptide
 US-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSY 10
 Db 1 DDIVPCMSY 10

RESULT 15

US-09-521-650-66
 ; Sequence 66, Application US/09521650
 ; Patent No. 6335429
 ; GENERAL INFORMATION:
 ; APPLICANT: Weber, Eckard
 ; APPLICANT: Cai, Sui Xiong
 ; APPLICANT: Keana, John F.W.
 ; APPLICANT: Drewe, John A.
 ; APPLICANT: Zhang, Han-Zhong
 ; TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
 ; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
 ; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
 ; TITLE OF INVENTION: Use Thereof
 ; FILE REFERENCE: 1735.0290002
 ; CURRENT APPLICATION NUMBER: US/09/521,650
 ; CURRENT FILING DATE: 2000-03-08
 ; EARLIER APPLICATION NUMBER: 09/168,888
 ; EARLIER FILING DATE: 1998-10-09
 ; EARLIER APPLICATION NUMBER: US 60/061,582
 ; EARLIER FILING DATE: 1997-10-10
 ; EARLIER APPLICATION NUMBER: US 09/033,661
 ; EARLIER FILING DATE: 1998-03-03
 ; NUMBER OF SEQ ID NOS: 142
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 66
 ; LENGTH: 10
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
 ; OTHER INFORMATION: Peptide
 US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVVPGMSY 10
 Db 1 DDIVPCMSY 10

Search completed: June 3, 2004, 12:03:07
 Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	96.2	11	12	US-09-909-164-5
2	50	96.2	11	12	US-09-909-164-6
3	50	96.2	11	12	US-09-909-164-9
4	50	96.2	11	12	US-09-909-164-10
5	50	96.2	11	12	US-09-909-164-47
6	50	96.2	11	12	US-09-909-164-48
7	50	96.2	11	12	US-09-909-164-49
8	50	96.2	11	12	US-09-909-164-50
9	50	96.2	11	12	US-09-909-164-51
10	50	96.2	11	12	US-09-909-164-52
11	46	88.5	11	12	US-09-909-164-8
12	46	88.5	11	12	US-09-909-164-12
13	46	88.5	11	12	US-09-909-164-13
14	46	88.5	11	12	US-09-909-164-7
15	45	86.5	11	12	US-09-909-164-11

16	45	86.5	11	12	US-09-909-164-19
17	45	86.5	11	12	US-09-909-164-20
18	45	86.5	11	12	US-09-909-164-23
19	45	86.5	11	12	US-09-909-164-24
20	44	84.6	11	12	US-09-909-164-28
21	44	84.6	11	12	US-09-909-164-29
22	44	84.6	11	12	US-09-909-164-33
23	44	84.6	11	12	US-09-909-164-36
24	44	84.6	11	12	US-09-909-164-37
25	44	84.6	11	12	US-09-909-164-43
26	42	80.8	11	12	US-09-909-164-14
27	41	78.8	11	12	US-09-909-164-22
28	41	78.8	11	12	US-09-909-164-26
29	41	78.8	11	12	US-09-909-164-27
30	41	78.8	11	12	US-09-909-164-61
31	41	78.8	11	12	US-09-909-164-62
32	40	76.9	11	12	US-09-909-164-21
33	40	76.9	11	12	US-09-909-164-25
34	40	76.9	11	12	US-09-909-164-31
35	40	76.9	11	12	US-09-909-164-32
36	40	76.9	11	12	US-09-909-164-35
37	40	76.9	11	12	US-09-909-164-40
38	40	76.9	11	12	US-09-909-164-41
39	40	76.9	11	12	US-09-909-164-45
40	40	76.9	11	12	US-09-909-164-46
41	39	75.0	11	12	US-09-909-164-30
42	39	75.0	11	12	US-09-909-164-34
43	39	75.0	11	12	US-09-909-164-38
44	39	75.0	11	12	US-09-909-164-39
45	39	75.0	11	12	US-09-909-164-42

ALIGNMENTS

RESULT 1

US-09-909-164-5
; Sequence, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-5

Query Match 96.2% Score 50; DB 12; Length 11;

```

Best Local Similarity 100.0%; Pred. No. 0.0014; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;

QY 1 BEVVPXGMSYS 11
   |||||
Db 1 BEVVPXGMSYS 11

RESULT 2
US-09-909-164-6
; Sequence 6, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-6
; Query Match 96.2%; Score 50; DB 12; Length 11;
; Best Local Similarity 100.0%; Pred. No. 0.0014;
; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
   |||||
Db 1 BEVVPXGMSYS 11

RESULT 3
US-09-909-164-9
; Sequence 9, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21

```

```

; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-9
; Query Match 96.2%; Score 50; DB 12; Length 11;
; Best Local Similarity 100.0%; Pred. No. 0.0014;
; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
   |||||
Db 1 BEVVPXGMSYS 11

RESULT 4
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)

```

OTHER INFORMATION: D-amino acids
US-09-909-164-10

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
| | | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 5

US-09-909-164-47
; Sequence 47, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: valine-(CO)
US-09-909-164-47

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
| | | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 6

US-09-909-164-48
; Sequence 48, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: leucine-(CO)
US-09-909-164-48

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
| | | | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 7

US-09-909-164-49
; Sequence 49, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norleucine-(CO)
US-09-909-164-49

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11

Db 1 EEVVPXGMSYS 11
|||||

RESULT 8

US-09-909-164-50
; Sequence 50, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: 2-amino-butyric acid-(CO)
US-09-909-164-50

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
|||||

Db 1 EEVVPXGMSYS 11
|||||

RESULT 9

US-09-909-164-51
; Sequence 51, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1

; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: (s,s)-allothreonine-(CO)
US-09-909-164-51

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
|||||

Db 1 EEVVPXGMSYS 11
|||||

RESULT 10

US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: propynyl glycine-(CO)
US-09-909-164-52

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
|||||

Db 1 EEVVPXGMSYS 11
|||||

RESULT 11

US-09-909-164-8
; Sequence 8, Application US/09909164
; Publication No. US20020068702A1

```
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
/ APPLICANT: Levy, Odile E
/ APPLICANT: Bruck, Terence K
/ TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
/ FILE REFERENCE: IN01192-US
/ CURRENT APPLICATION NUMBER: US/09/909,164
/ PRIOR FILING DATE: 2003-03-25
/ PRIOR APPLICATION NUMBER: 60/220,101
/ PRIOR FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 62
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 8
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: 11-mer synthesized according to example 1
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: ACETYLATION
/ NAME/KEY: MISC FEATURE
/ LOCATION: (6)..(6)
/ OTHER INFORMATION: norvaline-(CO)
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (9)..(9)
/ OTHER INFORMATION: D-amino acid
/ NAME/KEY: MOD RES
/ LOCATION: (11)..(11)
/ OTHER INFORMATION: AMIDATION
/ US-09-909-164-8
Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
      |||||
Db      1 BEVVPXGMDYS 11

RESULT 13
US-09-909-164-13
/ Sequence 13, Application US/09909164
/ Publication No. US20020068702A1
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
/ APPLICANT: Levy, Odile E
/ APPLICANT: Bruck, Terence K
/ TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
/ FILE REFERENCE: IN01192-US
/ CURRENT APPLICATION NUMBER: US/09/909,164
/ CURRENT FILING DATE: 2003-03-25
/ PRIOR APPLICATION NUMBER: 60/220,101
/ PRIOR FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 62
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 13
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: 11-mer synthesized according to example 1
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: ACETYLATION
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (11)..(11)
/ OTHER INFORMATION: AMIDATION
/ NAME/KEY: MISC FEATURE
/ LOCATION: (6)..(6)
/ OTHER INFORMATION: norvaline-(CO)
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (8)..(9)
/ OTHER INFORMATION: D-amino acids
/ US-09-909-164-13
Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
      |||||
Db      1 BEVVPXGMDYS 11

RESULT 12
US-09-909-164-12
/ Sequence 12, Application US/09909164
/ Publication No. US20020068702A1
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
/ APPLICANT: Levy, Odile E
/ APPLICANT: Bruck, Terence K
/ TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
/ FILE REFERENCE: IN01192-US
/ CURRENT APPLICATION NUMBER: US/09/909,164
/ CURRENT FILING DATE: 2003-03-25
/ PRIOR APPLICATION NUMBER: 60/220,101
/ PRIOR FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 62
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 12
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: 11-mer synthesized according to example 1
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: ACETYLATION
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (11)..(11)
/ OTHER INFORMATION: AMIDATION
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: ACETYLATION
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: ACETYLATION
/ US-09-909-164-9
```

```

RESULT 14
US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-7

```

```

Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 EEVVPXGMSYS 11
   |||||
Db 1 EEVVPXGMHYS 11

```

```

RESULT 15
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

```

```

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-11

```

```

Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 EEVVPXGMSYS 11
   |||||
Db 1 EEVVPXGMHYS 11

```

```

Search completed: June 3, 2004, 12:57:15
Job time : 33.7333 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 BEVFXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	T31308	hypothetical 367K
2	37	71.2	840	T39116	probable sulfate p
3	37	71.2	877	T40413	sulfate permease -
4	36	69.2	102	A42452	V1 protein - tobac
5	36	69.2	1498	B97355	DNA segregation AT
6	35	67.3	225	S57810	hypothetical prote
7	35	67.3	425	T24111	hypothetical prote
8	35	67.3	670	S22293	zinc finger protei
9	35	67.3	749	H82691	topoisomerase IV s
10	35	67.3	2717	A34203	DNA-binding protei
11	34	65.4	156	S54619	hypothetical prote
12	34	65.4	252	H69491	cell division inhi
13	34	65.4	544	C82900	probable ABC subst
14	33	63.5	94	T40758	hypothetical prote
15	33	63.5	116	E90544	50S ribosomal prot
16	33	63.5	165	D69493	hypothetical prote
17	33	63.5	253	C81374	hypothetical prote
18	33	63.5	259	T34536	hypothetical prote
19	33	63.5	284	S75817	beta-ketacyl-ACP
20	33	63.5	298	T47670	hypothetical prote
21	33	63.5	368	F72281	3-phosphoshikimate
22	33	63.5	426	D82163	cds37 protein - fi
23	33	63.5	466	T43653	iron(III) ABC tran
24	33	63.5	653	D82352	bacteriocin BCNS -
25	33	63.5	890	A30481	ATP-dependent DNA
26	33	63.5	1028	AF3286	conserved hypothet
27	33	63.5	1152	D87046	transforming growt
28	33	63.5	1394	A35626	DNA-directed RNA p
29	33	63.5	1401	G82336	

30	33	63.5	1548	2	T04456	hypothetical prote
31	33	63.5	1712	2	A38261	masking protein pr
32	32	61.5	84	2	E97333	hypothetical prote
33	32	61.5	175	2	P00616	transport protein
34	32	61.5	223	2	T01457	rho protein GDP-di
35	32	61.5	279	2	B72481	hypothetical prote
36	32	61.5	288	2	JC4011	cyclin D2 - rat
37	32	61.5	288	2	I58372	cyclin D2 - rat
38	32	61.5	289	2	A41984	cyclin D2 - mouse
39	32	61.5	289	2	A42842	cyclin D2 - human
40	32	61.5	291	2	S57922	cyclin D1 - Africa
41	32	61.5	291	2	S57925	cyclin D2 - chicke
42	32	61.5	291	2	JC4579	cyclin D1 - chicke
43	32	61.5	291	2	S62730	cyclin D1 - zebra
44	32	61.5	292	2	B42822	cyclin D3 - human
45	32	61.5	295	2	A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C/Species: Cenarchaeum symbiosum
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C/Accession: T31308
R/Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A/Title: Genomic analysis reveals chromosomal variation in natural populations of the u
A/Reference number: Z20994; MUID:98422450; PMID:9748430
A/Accession: T31308
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3472 <SCH>
A/Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC62699.1
C/Supfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;

Best Local Similarity 54.5%; Pred. No. 60;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVFXGMSYS 11

Db 2294 EDVIRGISFS 2304

RESULT 2

T39116

probable sulfate permease - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T39116

R/Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, November 1999

A/Reference number: Z21829

A/Accession: T39116

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-840 <HUN>

A/Cross-references: EMBL:AL132779; PIDN:CAB60015.1; GSPDB:GN000066; SPDB:SPAC869.05C

A/Experimental source: strain 972h-; cosmid c869

C/Genetics:

A/Gene: SPDB:SPAC869.05C

A/Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;

Best Local Similarity 77.8%; Pred. No. 21;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

Db 135 VVPQMSYA 143

```
RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Bartell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-877 <LYN>
A:Cross-references: EMBL:AL031261; PIDN:CAA30298.1; GSPDB:GNO0067; SPDB:SPBC3H7.02
A:Experimental source: strain 972h-; cosmid c3H7
C:Genetics:
A:Gene: SPDB:SPBC3H7.02
A:Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 148 VVPQMSYA 156

RESULT 4
A42452
VI protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
dwarf virus
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
DB 7 QVVPNGKTYA 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clos
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
stridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KUP>
A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:gl5026814; GSPDB:GNO0168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1276 EQKIPGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPP11) - tomato
C:Species: Lycopersicon esculentum (tomato)
C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of pistil-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MIL>
A:Cross-references: EMBL:U20592; NID:G924625; PIDN:AAA80497.1; PID:G924626
C:Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 32 DEVVPNGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Faricy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WIL>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GNO00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 335 EQIVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Wichelmeire, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991
```

A>Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
 A/Reference number: I58280; MUID:91187610; PMID:1901405
 A/Accession: S22293
 A>Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-670 <MIT>
 A/Cross-references: EMBL:X54250; NID:G57519; PIDN:CAA38151.1; PID:G57520
 A/Note: the authors did not translate the codon for residue 1
 C/Superfamily: HIV-EP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
 Best Local Similarity 66.7%; Pred. NO. 43;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 376 VVPAGLTYS 394

RESULT 9
 H82691
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C/Accession: H82691
 C/Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
 Nature 406, 151-157, 2000
 A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A82515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: H82691
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-749 <SIM>
 A/Cross-references: GB:AE003967; GB:AE003849; NID:G9106347; PIDN:AAF84162.1; GSPDB:GN001
 A/Experimental source: strain 945c
 R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.R.; Carraro, D.M.; Carver, H
 as-Neto, E.; Dorena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laigh
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.P.M.; Miracca, E.C.; Miyaki, C.V.
 ; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Genes: Xfl353
 C/Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase C

Query Match 67.3%; Score 35; DB 2; Length 749;
 Best Local Similarity 77.8%; Pred. NO. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 DB 526 EVDPSGMSY 534

RESULT 10
 A34203
 DNA-binding protein PRDII-BF1 - human
 N/Alternate names: major histocompatibility complex enhancer-binding protein 1
 C/Species: Homo sapiens (man)
 C/Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
 C/Accession: A34203; A34779
 R/Fan, C.M.; Maniatis, T.
 Genes Dev. 4, 29-42, 1990

A/Title: A DNA-binding protein containing two widely separated zinc finger motifs that
 A/Reference number: A34203; MUID:90169514; PMID:2106471
 A/Accession: A34203
 A>Status: Preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2717 <FAN>
 A/Cross-references: EMBL:X51435; NID:G38017; PIDN:CAA35798.1; PID:G38018
 R/Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
 Mol. Cell. Biol. 10, 1406-1414, 1990
 A/Title: A large protein containing zinc finger domains binds to related sequence eleme
 A/Reference number: A34779; MUID:90205817; PMID:2108316
 A/Accession: A34779
 A>Status: Preliminary; nucleic acid sequence not shown
 A/Residues: 801-1072; NID:1074-1168; NID:1170-1225; NID:1227-1434; NID:1436-1607; NID:1609-1
 A/Cross-references: GB:M32019
 C/Superfamily: HIV-EP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
 Best Local Similarity 66.7%; Pred. NO. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 2405 VVPAGLTYS 2413

RESULT 11
 S54619
 A/Title: The yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
 C/Species: Saccharomyces cerevisiae
 C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 A/Accession: S54619; S66879
 R/de Haan, M.; Maarse, A.C.; Grivell, L.A.
 submitted to the EMBL Data Library, May 1995
 A/Reference number: S54617
 A/Accession: S54619
 A/Molecule type: DNA
 A/Residues: 1-156 <DEH>
 A/Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123
 R/de Haan, M.; Grivell, L.A.; Maarse, A.C.
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: S66877
 A/Accession: S66879
 A/Molecule type: DNA
 A/Residues: 1-156 <DEH>
 A/Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA99201.1; PID:G1420111; MIFS:YORO
 A/Experimental source: strain S288C
 C/Genetics:
 A/Cross-references: SGD:S0005539
 A/Map position: 15R
 C/Superfamily: hypothetical protein YOR013W

Query Match 65.4%; Score 34; DB 2; Length 156;
 Best Local Similarity 66.7%; Pred. NO. 14;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 DB 50 EWPFLGMDY 58

RESULT 12
 H69491
 cell division inhibitor (minD-2) homolog - Archaeoglobus fulgidus
 C/Species: Archaeoglobus fulgidus
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 R/Accession: H69491
 R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
 ; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.P.; McDonald, L.

Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AE000970; GB:AE000782; MID:g2689293; PIDN:AAB89318.1; PID:g264860
 C:Superfamily: cell division inhibitor mind

Query Match 55.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVVPXGMS 9
 Db 81 EVIPAGMS 88

RESULT 13
 C82900
 probable ABC substrate-binding protein, iron UU359 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AF002133; GB:AF222894; MID:g6899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCebp-5; UU359
 A:Genetic code: GCG

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSY 10
 Db 135 EVVPHYLSY 144

RESULT 14
 I40758
 hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohyd
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:Z36940; MID:g535805; PIDN:CAA85392.1; PID:g535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVVPXGMSY 10
 Db 26 DIFFPGMSY 34

RESULT 15

E90544
 50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
 A:Reference number: A95512; MUID:21267165; PMID:11353084
 A:Accession: E90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:g14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPU_2610
 A:Genetic code: GCG
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
 Db 68 VRPLGMSYS 76

Search completed: June 3, 2004, 12:00:00
 Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds

(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	1058	1	Q8986 fusobacteri
2	37	71.2	877	1	O74377 schizosacch
3	36	69.2	102	1	P31619 tobacco yel
4	36	69.2	1498	1	Q04351 clostridium
5	35	67.3	2171	1	P15822 homo sapien
6	34	65.4	788	1	P32622 neurospora
7	34	65.4	1499	1	O60312 homo sapien
8	33	63.5	116	1	Q98qvo mycoplasma
9	33	63.5	165	1	O28330 archaeoglob
10	33	63.5	253	1	P45489 campylobact
11	33	63.5	280	1	Q9d387 mus musculu
12	33	63.5	426	1	Q9krb0 vibrio chol
13	33	63.5	466	1	O94740 schizosacch
14	33	63.5	478	1	Q9nm5 homo sapien
15	33	63.5	890	1	P08696 clostridium
16	33	63.5	1389	1	Q8cg18 mus musculu
17	33	63.5	1394	1	P22064 homo sapien
18	33	63.5	1401	1	Q9kv29 vibrio chol
19	33	63.5	1595	1	O14766 homo sapien
20	33	63.5	1712	1	Q00918 rattus norv
21	33	63.5	1713	1	Q8cg19 mus musculu
22	32	61.5	288	1	Q04827 rattus norv
23	32	61.5	289	1	P30279 homo sapien
24	32	61.5	289	1	P30280 mus musculu
25	32	61.5	291	1	Q90459 brachydanio
26	32	61.5	291	1	P49706 xenopus lae
27	32	61.5	291	1	P49706 gallus gall
28	32	61.5	291	1	P53782 xenopus lae
29	32	61.5	292	1	P55169 gallus gall
30	32	61.5	292	1	P30281 homo sapien
31	32	61.5	295	1	P24385 homo sapien
32	32	61.5	295	1	P25322 mus musculu
33	32	61.5	295	1	P39948 rattus norv

34	32	61.5	427	1	TOLB_HAEIN	P44677 haemophilus
35	32	61.5	529	1	ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1	S216_HUMAN	Q9y616 homo sapien
37	32	61.5	719	1	GSP_CRIFA	P90518 crithidia f
38	32	61.5	726	1	PRTE_HSV6U	P52384 human herpe
39	32	61.5	726	1	PRTP_HSV6Z	P52544 human herpe
40	32	61.5	759	1	SCTI_YEAST	P32784 saccharomyc
41	32	61.5	920	1	RDD_RAT	Q62671 rattus norv
42	32	61.5	993	1	VIA_TAV	P28891 tomato aspe
43	32	61.5	1377	1	RHSA_ECOLI	P16916 escherichia
44	32	61.5	1397	1	RHSC_ECOLI	P16918 escherichia
45	32	61.5	1411	1	RHSE_ECOLI	P16917 escherichia

ALIGNMENTS

RESULT 1						
CARB_FUSNN						
ID	CARB_FUSNN	STANDARD;	PRT;	1058	AA.	
AC	Q8986;					
DT	28-FEB-2003	(Rel. 41, Created)				
DT	28-FEB-2003	(Rel. 41, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).					
DE	phosphate synthetase ammonia chain).					
GN	CARB OR FN0422.					
OS	Fusobacterium nucleatum (subsp. nucleatum).					
OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;					
OC	Fusobacterium.					
OX	NCBI_TaxID=76856;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATCC 25586;					
RX	MEDLINE=21886394; PubMed=1189109;					
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,					
RA	Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,					
RA	Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,					
RA	Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,					
RA	Fonstein M., Kyripides N., Overbeek R.;					
RT	"Genome sequence and analysis of the oral bacterium Fusobacterium					
RT	nucleatum strain ATCC 25586."					
RL	J. Bacteriol. 184:2005-2018 (2002).					
CC	-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +					
CC	phosphate + L-glutamate + carbamoyl phosphate.					
CC	-!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).					
CC	-!- PATHWAY: Arginine biosynthesis.					
CC	-!- PATHWAY: Pyrimidine biosynthesis; first step.					
CC	-!- SUBUNIT: Composed of two chains; the small (or glutamine) chain					
CC	promotes the hydrolysis of glutamine to ammonia, which is used by					
CC	the large (or ammonia) chain to synthesize carbamoyl phosphate (By					
CC	similarity).					
CC	-!- SIMILARITY: Belongs to the carb family.					
CC						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC						
CC	EMBL; A010554; AAL94625.1; ALT_INIT.					
DR	HAWAF; MF_01210; ; 1.					
DR	InterPro; IPR006275; CarA_L_glu.					
DR	InterPro; IPR005483; CPase_L.					
DR	InterPro; IPR005479; CPase_L_D2.					
DR	InterPro; IPR005480; CPase_L_D3.					
DR	InterPro; IPR005481; CPase_L_N.					
DR	InterPro; IPR004362; MGS-like.					
DR	Pfam; PF00289; CPasease_L_chain; 2.					
DR	Pfam; PF02786; CPasease_L_D2; 2.					
DR	Pfam; PF02787; CPasease_L_D3; 1.					


```

DR PFAM: PF02142; MGS: 1.
DR PRINTS: PF00098; CPSASE
DR TIGRFAMs: TIGR01369; CPSASE1; 1.
DR PROSITE: PS00866; CPSASE_1; 2.
DR PROSITE: PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 284
FT METAL 298 298
FT METAL 300 300
FT METAL 820 820
FT METAL 832 832
FT METAL 832 832
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF7C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVFPGMSYS 11
DB 190 EIVPGLNYS 199
|:|:| |:|:|
|:|:| |:|:|

RESULT 2
SULH_SCHPO
ID SULH_SCHPO STANDARD; PRT; 877 AA.
AC 074377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable sulfate permease C3H7.02.
GN SPBC3H7.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Murgall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard J.C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez C., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McComb W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53)
CC family.
CC -1- SIMILARITY: Contains 1 STAS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL031261; CAA20298.1; -.
CC PIR: T40413; T40413.
CC GeneDB_Sfombe; SPBC3H7.02; -.
CC InterPro: IPR002645; STAS.
CC InterPro: IPR001902; Sulph_transp.
CC Pfam: PF01740; STAS; 1.
CC Pfam: PF00916; Sulfate transp; 1.
CC TIGRFAMs: TIGR00815; sulp; 1.
CC PROSITE: PS01130; SLC26A; 1.
CC PROSITE: PS08081; STAS; 1.
CC Transports; transmembrane.
CC TRANSMEM 133 153 POTENTIAL.
CC TRANSMEM 161 181 POTENTIAL.
CC TRANSMEM 186 206 POTENTIAL.
CC TRANSMEM 221 241 POTENTIAL.
CC TRANSMEM 243 263 POTENTIAL.
CC TRANSMEM 292 312 POTENTIAL.
CC TRANSMEM 329 349 POTENTIAL.
CC TRANSMEM 384 404 POTENTIAL.
CC TRANSMEM 424 444 POTENTIAL.
CC TRANSMEM 461 481 POTENTIAL.
CC TRANSMEM 484 504 POTENTIAL.
CC TRANSMEM 518 538 POTENTIAL.
CC TRANSMEM 543 563 POTENTIAL.
CC DOMAIN 594 747 STAS
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
Best Local Similarity 77.8%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
DB 148 VVFGMSYA 156
|:|:| |:|:|
|:|:| |:|:|

RESULT 3
V1LK_TYDVA
ID V1LK_TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Hailey A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81103; AAA47947.1; -.
DR PIR; A42452; A42452.
DR InterPro; IPR002621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; I.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 1.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 7 QVVPXGINS 16
:|||||:|
:

RESULT 4
YL9S_CLOAB STANDARD; PRT; 1498 AA.
ID YL9S_CLOAB
AC Q04351;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein CAC3709.
GN CAC3709
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE=21359325; PubMed=11466286;
RX Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatouev R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
RN [2]
RP SEQUENCE OF 1-108 FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE=93273706; PubMed=8501044;
RX Sauer U., Duerre P.;
RT "Sequence and molecular characterization of a DNA region encoding a
RT small heat shock protein of Clostridium acetobutylicum.";
RL J. Bacteriol. 175:3394-3400(1993).
CC -1- SIMILARITY: Contains 2 FtsK domains.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in positions 76 and 106.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007866; AX81629.1; -.
DR EMBL; X65276; CAA46379.1; ALT_FRAME.
DR PIR; B97355; B97355.
DR InterPro; IPR002543; FtsK_SpoIIIE.

DR Pfam; PF01580; FtsK_SpoIIIE; 2.
DR PROSITE; PS50901; FTSK; 2.
KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
FT DOMAIN 655 857 FTSK 1.
FT DOMAIN 1001 1188 FTSK 2.
FT NP_BIND 675 682 ATP (POTENTIAL).
SQ SEQUENCE 1498 AA; 168968 MW; FF42037A35A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
Db 1276 EQKIPMGMSY 1285
:|||||:|
:

RESULT 5
ZEPI_HUMAN STANDARD; PRT; 2717 AA.
ID ZEPI_HUMAN
AC P15922;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
DE binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE (PRDII-BF1).
GN HIVEP1 OR ZNF40.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90168514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger
RT motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
RA Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger
RT from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142
RX MEDLINE=92232684; PubMed=1567844;
RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: By mitogens and phorbol ester.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVEP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL: X51435; CAA35798.1; -
 PIR: A34203; A34203.
 DR PDB; 3ZNF; 15-JAN-92.
 DR PDB; 4ZNP; 15-JAN-92.
 DR PDB; 1BBO; 31-OCT-93.
 DR TRANSFAC; T00497; -
 DR Genew; HGNC:4920; HIVEP1.
 DR MIM; 194540; -
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003877; F:DNA binding; TAS.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00936; zf-C2H2; 5.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Repeat; 3D-structure.
 FT ZN_FING 406 428
 FT ZN_FING 434 456
 FT ZN_FING 958 981
 FT ZN_FING 2087 2109
 FT ZN_FING 2115 2139
 FT ZN_FING 803 806
 FT DOMAIN 2088 2088
 FT STRAND 2090 2092
 FT TURN 2095 2095
 FT STRAND 2109 2109
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. NO. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
 ||| |::|
 Db 2405 VVPAGLTYS 2413

RESULT 6

ID CY14 NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate permease II.
 GN CYS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129256; PubMed=1825178;
 RA Ketter J.S., Jara G., Fu Y.-H., Warzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of cys-14, the structural gene for sulfate permease II in
 RT Neurospora crassa.";
 RL Biochemistry 30:1780-1787(1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RX MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate

permease II and a putative human tumour suppressor.";
 Trends Biochem Sci 19:19-19(1994).
 CC -!- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- INDUCTION: Highly expressed, but only in cells subject to sulfur
 CC limitation, and it is turned on by the positive-acting Cys-3
 CC sulfur regulatory protein.
 CC -!- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC -!- SIMILARITY: Belongs to the SLC26A/sulph transporter (TC 2.A.53)
 CC family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: M59167; AAA33615.1; ALT_SEQ.

PIR: A37956; A37956. Sulph_transpt.

InterPro; IPR001902; Sulph_transpt.

Pfam; PF00916; Sulfate_transp; 1.

TIGRFAMS; TIGR00815; sulp; 1.

PROSITE; PS01130; SLC26A; 1.

Transport; Transmembrane; Glycoprotein.

TRANSMEM 71 91 POTENTIAL.

TRANSMEM 103 123 POTENTIAL.

TRANSMEM 128 148 POTENTIAL.

TRANSMEM 171 191 POTENTIAL.

TRANSMEM 193 213 POTENTIAL.

TRANSMEM 271 291 POTENTIAL.

TRANSMEM 326 346 POTENTIAL.

TRANSMEM 363 383 POTENTIAL.

TRANSMEM 451 471 POTENTIAL.

TRANSMEM 474 494 POTENTIAL.

CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. NO. 36;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
 ||| |::|
 Db 90 VVPGMAYA 98

RESULT 7

ID A10A HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; Q96914;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
 DE (Aminophospholipid translocase VA).
 GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21225279; PubMed=11326269;
 RA Meguro M., Kashiwagi A., Mitsuura K., Nakao M., Kondo I., Saitoh S.,
 RA Oshimura M.;
 RT "A novel maternally expressed gene, ATP10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome.";
 RL Nat. Genet. 28:19-20(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

XX MEDLINE=21313119; PubMed=11353404;
 RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
 RT "The human aminophospholipid-translocating ATPase gene ATP10C maps
 adjacent to UBE3A and exhibits similar imprinted expression.";
 RL Am. J. Hum. Genet. 58:1501-1505(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 337-1499 FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- TISSUE SPECIFICITY: Widely expressed, with highest levels in
 kidney, followed by lung, brain, prostate, testis, ovary and
 small intestine.
 CC -!- DISEASE: Defects in ATP10A are a cause of Angelman syndrome (AS)
 [MIM:105830]; also known as 'happy puppet syndrome'. AS is
 characterized by features of severe motor and intellectual
 retardation, microcephaly, ataxia, frequent jerky limb movements
 and flapping of the arms and hands, hypotonia, hyperactivity,
 hypopigmentation, seizures, absence of speech, frequent smiling
 and episodes of paroxysmal laughter, and an unusual facies
 characterized by macrostomia, a large mandible and open-mouthed
 expression, a great propensity for protruding the tongue ('tongue
 thrusting'), and an occipital groove.
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 ATPases). Subfamily IV.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; AB051358; BAB47392.1; -
 CC EMBL; AY029504; AAK33100.1; -
 CC EMBL; AY029487; AAK33100.1; JOINED.
 CC EMBL; AY029488; AAK33100.1; JOINED.
 CC EMBL; AY029489; AAK33100.1; JOINED.
 CC EMBL; AY029490; AAK33100.1; JOINED.
 CC EMBL; AY029491; AAK33100.1; JOINED.
 CC EMBL; AY029492; AAK33100.1; JOINED.

DR EMBL; AY029493; AAK33100.1; JOINED.
 DR EMBL; AY029494; AAK33100.1; JOINED.
 DR EMBL; AY029495; AAK33100.1; JOINED.
 DR EMBL; AY029496; AAK33100.1; JOINED.
 DR EMBL; AY029497; AAK33100.1; JOINED.
 DR EMBL; AY029498; AAK33100.1; JOINED.
 DR EMBL; AY029499; AAK33100.1; JOINED.
 DR EMBL; AY029500; AAK33100.1; JOINED.
 DR EMBL; AY029501; AAK33100.1; JOINED.
 DR EMBL; AY029502; AAK33100.1; JOINED.
 DR EMBL; AY029503; AAK33100.1; JOINED.
 DR EMBL; BC052351; AAK52251.1; -
 DR EMBL; AB011138; BAB25492.1; -
 DR Genew; HGNC:13542; ATP10A.
 DR MIM; 605855; -
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0004012; P:phospholipid-translocating ATPase activity; NAS.
 DR GO; GO:0008360; P:regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase_E1-E2.
 DR InterPro; IPR006539; Flippase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type; 6.
 DR PROSITE; PS00154; ATPase_E1-E2; 1.
 DR Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Multigene family.
 FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 87 106 POTENTIAL.
 FT DOMAIN 107 110 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 111 128 POTENTIAL.
 FT DOMAIN 129 309 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 310 332 POTENTIAL.
 FT DOMAIN 333 362 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 363 384 POTENTIAL.
 FT DOMAIN 385 1087 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1088 1108 POTENTIAL.
 FT DOMAIN 1109 1119 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1120 1140 POTENTIAL.
 FT DOMAIN 1141 1170 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1171 1192 POTENTIAL.
 FT DOMAIN 1193 1199 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1200 1222 POTENTIAL.
 FT DOMAIN 1223 1228 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1229 1249 POTENTIAL.
 FT DOMAIN 1250 1267 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1268 1292 POTENTIAL.
 FT DOMAIN 1293 1499 CYTOPLASMIC (POTENTIAL).
 FT MOD RES 427 427 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 1031 1031 MAGNESIUM (BY SIMILARITY).
 FT METAL 1035 1035 MAGNESIUM (BY SIMILARITY).
 FT DOMAIN 467 470 POLY-GLU.
 FT CONFLICT 388 388 O -> R (IN REF. 4).
 SQ SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 70;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGMSYS 11
 DB 469 EEVVPXGMSYS 479
 RESULT 8
 RL20 MYCPU
 ID RL20 MYCPU STANDARD; PRT; 116 AA.
 AC Q98QV0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 50S ribosomal protein L20.
GN RPLT OR MYPU 2610.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
CC necessary for the in vitro assembly process of the 50s ribosomal
CC subunit. It is not involved in the protein synthesizing functions
CC of that subunit (by similarity).
CC -1- SIMILARITY: Belongs to the L20p family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: A1445563; CAC13434.1; -;
DR F1R; E90544; E90544.
DR Mypulist; M1PU_2610; -;
DR HAMAP; MF_00382; -; 1.
DR InterPro; IPR005813; Ribosomal L20.
DR InterPro; IPR005812; Ribosomal L20b/o.
DR Pfam; PF00453; Ribosomal L20; 1.
DR PRINTS; PR00062; RIBOSOMALL20.
DR ProDom; PD002359; L20; 1.
DR TIGRFAMs; TIGR01032; Zpt1 bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 8.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 VTPXGMSYS 11
DB 68 VRPLGMSYS 76
RESULT 9
ID VJ49 ARCFU STANDARD; PRT; 165 AA.
AC Q28350; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.X., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uitterlind L.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000968; AA899307.1; -;
DR F1R; D59493; D69493.
DR TIGR; AF1949; -;
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 165 AA; 17588 MW; B8C17054810ADB88 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EVVFXGMSY 10
DB 60 EESIPDGASY 69
RESULT 10
ID Y990 CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990c.
GN Cj0990c.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jørgensen K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream A.M., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglutamine amidohydrolase (Hippuricase) gene in Escherichia
RT coli.";
RJ J. Bacteriol. 177:2396-2402(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@ieb-sib.ch)

DR EMBL; AL139076; CAB72246.1; --
 DR EMBL; Z36940; CAAB5392.1; --
 DR PIR; C81374; C81374.
 DR PIR; I40758; I40758.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
 Best Local Similarity 55.6%; Pred. No. 18; Indels 0; Gaps 0;
 Matches 5; Conservative 2; Mismatches 2;
 QY 2 EVVPGXMSY 10
 Db 185 DIPPGXMSY 193

RESULT 11

CTX3 MOUSE STANDARD; PRT; 280 AA.
 AC Q9D387; Q9CKQ4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 GN C20ORF103.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata Y., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690 (2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gurnatne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; AK014127; BAB29169.1; -
 DR EMBL; AK016222; BAB31124.1; ALT_FRAME.
 DR EMBL; EC004791; AAH04791.1; -
 DR MGD; MGI:1920368; 3110035N03Rik.
 DR MGD; MGI:1923411; 6330527O06Rik.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 280 PROTEIN C20ORF103 HOMOLOG.
 FT DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 236 256 POTENTIAL.
 FT DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1 118 Missing (in isoform 2).
 FT /FTID=VSP_003820
 FT CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).
 FT CONFLICT 230 230 Q -> P (IN REF. 1; BAB31124).
 FT CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).
 SQ SEQUENCE 280 AA; 31721 MW; FA11D7BF9FD5CCCF CRC64;

Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPGXMSY 10
 Db 173 VTPAGMSY 180

RESULT 12

AROA_VIBCH STANDARD; PRT; 426 AA.
 ID AROA_VIBCH
 AC Q9KRE0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPRS synthase) (EPRS).
 GN AROA OR VC1732.
 OS Vibrio cholerae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Urtreback T., Fleischmann R.D., Niernan W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RI "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RI cholerae";
 RL Nature 405:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE004251; AAF94882.1; -;
 DR PIR; D82163; D82163.
 DR TIGR; VC1732; -;
 DR HAVAP; MF 00210; -;
 DR InterPro; IPR006264; AroA.
 DR InterPro; IPR001986; EPSP_synth.
 DR Pfam; PF00275; EPSP_synthase; 1.
 DR ProDom; PD001867; EPSP_synthase; 1.
 DR TIGRFAMs; TIGR01356; aroA; 1.
 DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
 DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
 KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 426 AA; 45101 MW; 38852D6483BF1C3 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 426;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVFXGMSY 10
 Db 223 EFVIPAQGSY 232
 RESULT 13
 ID CC37 SCHPO STANDARD; PRT; 466 AA.
 AC 094740;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
 DE subunit) (Cell division control protein 37).
 GN CDC37 OR SPBC9B6.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Westwood P.K., Preston N.C., Fantes P.A.;
 RI "Schizosaccharomycetes pombe cdc37 gene";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Wordley J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weetjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert P., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.D., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT The genome sequence of Schizosaccharomycetes pombe.;
 RL Nature 415:871-880(2002).
 RN [3]
 RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=22745456; PubMed=12861001;
 RA Tatebe H., Shiozaki K.;
 RT "Identification of cdc37 as a novel regulator of the stress-responsive
 RT mitogen-activated protein kinase.";
 RL Mol. Cell. Biol. 23:5132-5142(2003).
 CC -1- FUNCTION: Co-chaperone that binds to numerous kinases and promotes
 CC their interaction with the Hsp90 complex, resulting in
 CC stabilization and promotion of their activity.
 CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with sty1.
 CC -1- SUBCELLULAR LOCATION: Nuclear, and cytoplasmic. When in the
 CC nucleus associated with chromatin.
 CC -1- SIMILARITY: Belongs to the CDC37 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ132377; CAB38758.1; -;
 DR EMBL; AJ132376; CAB38757.1; -;
 DR EMBL; AL049769; CAB42371.2; -;
 DR PIR; T43653; T43653.
 DR GeneDB SPombe; SPBC9B6.10; -;
 DR InterPro; IPR004918; Cdc37.
 DR Pfam; PF03234; Cdc37; 1.
 KW Chaperone; Cell division; Cell cycle; Nuclear protein.
 SQ SEQUENCE 466 AA; 52554 MW; 647238B34CAB3C5 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 466;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVFXGMSY 10
 Db 98 DSAIFGMSY 107
 RESULT 14


```
RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene."
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50,
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro."
RL Mol. Microbiol. 2:607-614(1988).
CC -!- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -!- INDUCTION: By UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -.
DR EMBL; M32882; AAA98249.1; -.
DR PIR; A30481; A30481.
DR InterPro; IPR000834; Peptidase_M14.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR SMART; SM00287; SH3b; 3.
KW Antibiotic; Bacteriocin; Plasmid.
FT DOMAIN 815 869 HYDROPHOBIC.
SQ SEQUENCE 890 AA; 96699 MW; F4E5E871C31C6C6 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred.No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 170 EWVPGGFTY 178
|||||
|::|
```

Search completed: June 3, 2004, 11:49:53
Job time : 5.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-9
Perfect score: 52
Sequence: 1 BEVFXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	1044	16 Q8DIH0	Q8diH0 synechococ
2	38	73.1	344	16 Q815A7	Q815A7 bacillus ce
3	38	73.1	387	16 Q8FX1	Q8fx1 rhizobium 1
4	38	73.1	3472	1 O74056	O74056 cenarchaeum
5	27	71.2	840	3 Q8URY8	Q8ury8 schizosacch
6	37	71.2	1123	16 Q8EW04	Q8ewd4 mycoplasma
7	36	69.2	471	11 Q8RL126	Q8rl126 mus musculu
8	36	69.2	484	11 Q8VD18	Q8vd18 mus musculu
9	36	69.2	484	11 Q8BTX4	Q8btx4 mus musculu
10	36	69.2	484	11 Q8BK35	Q8bk35 mus musculu
11	36	69.2	559	16 Q839T9	Q839t9 enterococcu
12	36	69.2	1399	16 Q889X7	Q889x7 pseudomonas
13	35	67.3	225	10 Q40129	Q40129 lycopersico
14	35	67.3	245	16 Q7V604	Q7v6q4 prochloroco
15	35	67.3	425	5 Q9XVK4	Q9xvk4 caenorhabdi
16	35	67.3	495	11 Q8C1D7	Q8cid7 mus musculu

17	35	67.3	555	4 Q7Z6R0	Q7z6r0 homo sapien
18	35	67.3	583	5 Q9BH83	Q9bh83 plasmodium
19	35	67.3	583	5 Q9BHA5	Q9bha5 plasmodium
20	35	67.3	583	5 Q815S7	Q815s7 plasmodium
21	35	67.3	670	11 Q01487	Q01487 rattus ratt
22	35	67.3	747	16 Q8PMI6	Q8pmi6 xanthomonas
23	35	67.3	747	16 Q8PAT2	Q8pat2 xanthomonas
24	35	67.3	749	16 Q9PDM6	Q9pdm6 xyella fas
25	35	67.3	1902	4 Q14122	Q14122 homo sapien
26	34	65.4	156	3 Q12479	Q12479 saccharomyc
27	34	65.4	219	17 Q971S2	Q971s2 sulfolobus
28	34	65.4	252	17 Q28342	Q28342 archaeoglob
29	34	65.4	290	4 Q96MU1	Q96mu1 homo sapien
30	34	65.4	387	16 Q92MD6	Q92md6 rhizobium m
31	34	65.4	489	4 Q81YM3	Q81ym3 homo sapien
32	34	65.4	541	16 Q98BP5	Q98bp5 rhizobium 1
33	34	65.4	544	16 Q9QD2	Q9qdd2 ureaplasma
34	34	65.4	731	16 Q7UWU7	Q7uwu7 rhodopirell
35	34	65.4	842	3 Q9URR4	Q9urra4 penicillium
36	34	65.4	899	16 Q8G4I5	Q8g4i5 bifidobacte
37	34	65.4	1049	16 Q8XT05	Q8xt05 ralstonia s
38	34	65.4	1400	16 Q8TKQ5	Q8tkq5 vibrio para
39	33	63.5	143	17 Q8TX62	Q8tx62 methanopyru
40	33	63.5	166	16 Q8PPP5	Q8ppp5 xanthomonas
41	33	63.5	193	2 Q8VJA8	Q8vua8 lactococcus
42	33	63.5	208	2 Q8KTQ4	Q8ktq4 candidatus
43	33	63.5	209	16 Q8RE56	Q8re56 fusobacteri
44	33	63.5	251	13 Q7SY67	Q7sy67 xenopus lae
45	33	63.5	282	16 Q7U552	Q7u552 synechococc

ALIGNMENTS

RESULT 1

ID	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
AC	Q8DIH0;		
DT	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)		
DE	Multidrug efflux transporter.		
GN	TLL1618.		
OS	Synechococcus elongatus (Thermosynechococcus elongatus).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI TaxID=32046;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-1;		
RX	MEDLINE=2225144; PubMed=12240834;		
RA	Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katch H., Sasamoto S.,		
RA	Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,		
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of the thermophilic cyanobacterium		
RT	Thermosynechococcus elongatus BP-1.";		
RL	DNA Res. 3:123-130(2002).		
DR	EMBL; AP005374; BAC09170.1; -		
DR	GO; GO:0015021; C:integral to membrane; IEA.		
DR	GO; GO:0005215; P:transporter activity; IEA.		
DR	GO; GO:0006810; P:transport; IEA.		
DR	InterPro; IPR001036; Acrflavin_res.		
DR	InterPro; IPR004764; HAE1		
DR	Pfam; PF00873; ACR_tran; 1.		
DR	PRINTS; PR00702; ACRIFLAVINRP.		
DR	TIGRFAMs; TIGR00915; 2A0602; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;		

Query Match 75.0%; Score 39; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 BEVVPXGMSYS 11
DB 843 BEVLPNGIGYS 853

RESULT 2
Q815A7 PRELIMINARY; PRT; 344 AA.
AC Q815A7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BCS259.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A. PubMed=12721630;
RX MEDLINE=24608415;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Hasselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017015; AAP12123.1; -.
DR InterPro; IPR000437; Prok lipoprot S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 344 AA; 38539 MW; C55268ACB7225995 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 344;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 10
DB 152 EBIAPLGUSY 161

RESULT 3
Q98FX1 PRELIMINARY; PRT; 387 AA.
AC Q98FX1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hippurate hydrolase.
GN MLR3583.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AF003002; BAB5045.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metalloproteinase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase_M20.

DR Pfam; PF01546; Peptidase_M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 10
DB 367 DEAIPIHGMYS 376

RESULT 4
O74056 PRELIMINARY; PRT; 3472 AA.
AC O74056;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
OC Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B.
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL; AF083072; AAC52699.1; -.
DR PIR; T31308; T31308.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR001680; WD40.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
KW Hypothetical protein.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 2294 EDVIPRGISFS 2304

RESULT 5
Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972h-;
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;

```

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALI32779; CAB60015.1; -
DR PIR; T39116; T39116.
DR GenedB.SPombe; SPAC869.05c; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008271; P:sulfate porter activity; IEA.
DR GO; GO:0008272; P:sulfate transport; IEA.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transpt.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRFAMs; TIGR00815; sulp; 1.
DR PROSITE; P350801; STAS; 1.
SQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
||| |||||
DB 135 VVPGMSYA 143

RESULT 6
Q8EWD4 PRELIMINARY; PRT; 1123 AA.
AC Q8EWD4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MYPE 2560 paralog, 57%.
GN MYPE2710.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HF-2;
RX MEDLINE=22354719; PubMed12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004171; BAC44062.1; -
DR InterPro; IPR008985; ConA like lec gl.
DR InterPro; IPR007326; Lipoprotein_17.
DR Pfam; PF04200; Lipoprotein_17; 3.
KW Complete proteome.
SQ SEQUENCE 1123 AA; 123636 MW; A4D707330E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
Best Local Similarity 70.0%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVFXGMSYS 11
||| |||||
DB 658 EYVPMGLSYS 667

RESULT 7
Q8R126 PRELIMINARY; PRT; 471 AA.
AC Q8R126;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025810; AAH25810.1; -
DR MGD; MGI:2154441; Gltscr2.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;
Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVFXGMSYS 11
||| |||||
DB 226 EYVPGASYN 235

RESULT 8
Q8VD18 PRELIMINARY; PRT; 484 AA.
AC Q8VD18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AWS36441.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017637; AAH17637.1; -
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVFXGMSYS 11
||| |||||
DB 239 EYVPGASYN 248

RESULT 9
Q8BTX4 PRELIMINARY; PRT; 484 AA.
AC Q8BTX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.
RL Nature 420:563-573 (2002).
DR EMBL; AK088461; BAC40367.1; -.
DR MGD; MGI:2154441; Gtscr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5ECAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

QY 2 EVVPXGMSYS 11
|||:|:|:
DB 239 EVIPAGASYN 248

RESULT 10

Q8BK35 PRELIMINARY; PRT; 484 AA.

AC Q8BK35;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE Protein.
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Placental;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK077341; BAC36760.1; -.
DR MGD; MGI:2154441; Gtscr2.
SQ SEQUENCE 484 AA; 55792 MW; EB67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

QY 2 EVVPXGMSYS 11
|||:|:|:
DB 239 EVIPAGASYN 248

RESULT 11

Q839T9 PRELIMINARY; PRT; 559 AA.

AC Q839T9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pheromone binding protein, putative.
GN EF0063.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]_TaxID=1351;
RP SEQUENCE FROM N.A.

RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty J., DeBoy R.J., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;

RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074 (2003).
DR EMBL; AE016947; AAC79943.1; -.
DR TIGR; EF0063; -.
DR GO; GO:0005215; P:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok Lipoprot_S.
DR InterPro; IPR000914; SBB_bac_5.
DR Pfam; PF00496; SBB_bac_5_1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 559 AA; 61476 MW; CCI5418D3D53DE7 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 559;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
:|:|:|:|:
DB 351 LIPEGMSYS 359

RESULT 12

Q889X7 PRELIMINARY; PRT; 1399 AA.

AC Q889X7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA-directed RNA polymerase, beta' subunit.
GN RPOC OR PSPTO0620.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]_TaxID=323;
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldlyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen I.,
RA White O., Fraser C., Collmer A.;
RA "Complete sequence of Pseudomonas syringae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016858; AAC54162.1; -.
DR TIGR; PSPTO0620; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04999; RNA_pol_Rpb1_5; 1.
KW DNA-directed RNA polymerase; Complete proteome.
SQ SEQUENCE 1399 AA; 154733 MW; 26178BD653102ADE CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1399;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 10
:|:|:|:|:
DB 581 QVVPAGLSY 589

RESULT 13

Q40129 Q40129 PRELIMINARY; PRT; 225 AA.
 AC Q40129;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein precursor.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF36; TISSUE=stilt;
 RX MEDLINE=95375233; PubMed=7647301;
 RA Milligan S.B., Gasser C.S.;
 RT "Nature and regulation of pistil-expressed genes in tomato."
 RL Plant Mol. Biol. 28:691-711(1995).
 DR EMBL; U20592; AAA80497.1; -.
 DR PIR; S57810; S57810.
 DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.
 DR InterPro; IPR002160; Kunitz_legume.
 DR Pfam; PF00197; Kunitz_legume; 1.
 DR PRINTS; PR00291; KUNITZINHTR.
 DR ProDom; PD000891; Kunitz_legume; 1.
 DR SMART; SM00452; STI; 1.
 DR PROSITE; PS00283; SOYBEAN KUNITZ; 1.
 KW Hypothetical protein; Signal. POTENTIAL.
 FT SIGNAL 1 20
 FT CHAIN 21 225 UNKNOWN.
 SQ SEQUENCE 225 AA; 25188 MW; 1074C361D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
 Best Local Similarity 54.5%; Pred. No. 37;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
 Db 32 DEVVPNGKTYA 42

RESULT 14

Q7V6Q4 Q7V6Q4 PRELIMINARY; PRT; 245 AA.
 AC Q7V6Q4;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Phospholipid and glycerol acyltransferase (From 'motifs_6.mef').
 GN PWT1092.
 OS Prochlorococcus marinus (strain MT 9313).
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=74547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22825698; PubMed=12917642;
 RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
 RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
 RA Webb B.A., Zinser E.R., Chisholm S.W.;
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
 niche differentiation."
 RL Nature 424:1042-1047(2003).
 DR EMBL; BX572098; CAE21267.1; -.
 KW Acyltransferase; Transferase; Complete proteome.
 SQ SEQUENCE 245 AA; 26907 MW; 105F704CBE2C6427 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 245;

Best Local Similarity 60.0%; Pred. No. 41;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
 Db 179 QVVPVGLGYS 188

RESULT 15

Q9XVK4 Q9XVK4 PRELIMINARY; PRT; 425 AA.
 ID Q9XVK4;
 AC Q9XVK4;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE R10D12.10 protein.
 GN R10D12.10
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Percy C.M.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology".
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81109; CAB03241.1; -.
 DR PIR; T24111; T24111.
 DR WormPep; R10D12.10; CEI2690.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004672; F:protein kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR000719; Prot_Kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
 Best Local Similarity 50.0%; Pred. No. 75;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
 Db 335 EQIVPGLQY 344

Search completed: June 3, 2004, 11:57:32
 Job time : 29.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseq_29Jan04.*
- 2: Geneseq_1980s.*
- 3: Geneseq_1990s.*
- 4: Geneseq_2000s.*
- 5: Geneseq_2001s.*
- 6: Geneseq_2002s.*
- 7: Geneseq_2003s.*
- 8: Geneseq_2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	96.2	11	5	ABB80525 Hepatitis
2	50	96.2	11	5	ABB80521 Hepatitis
3	50	96.2	11	5	ABB80522 Hepatitis
4	50	96.2	11	5	ABB80566 Hepatitis
5	50	96.2	11	5	ABB80563 Hepatitis
6	50	96.2	11	5	ABB80565 Hepatitis
7	50	96.2	11	5	ABB80567 Hepatitis
8	50	96.2	11	5	ABB80559 Hepatitis
9	50	96.2	11	5	ABB80526 Hepatitis
10	50	96.2	11	5	ABB80564 Hepatitis
11	50	96.2	11	5	ABB80568 Hepatitis
12	46	88.5	11	5	ABB80561 Hepatitis
13	46	88.5	11	5	ABB80524 Hepatitis
14	46	88.5	11	5	ABB80529 Hepatitis
15	46	88.5	11	5	ABB80528 Hepatitis
16	46	88.5	11	5	ABB80562 Hepatitis
17	45	86.5	11	5	ABB80523 Hepatitis
18	45	86.5	11	5	ABB80536 Hepatitis
19	45	86.5	11	5	ABB80558 Hepatitis
20	45	86.5	11	5	ABB80560 Hepatitis
21	45	86.5	11	5	ABB80527 Hepatitis
22	45	86.5	11	5	ABB80535 Hepatitis
23	45	86.5	11	5	ABB80540 Hepatitis
24	45	86.5	11	5	ABB80539 Hepatitis
25	44	84.6	11	5	ABB80549 Hepatitis

26	44	84.6	11	5	ABB80544	Abb80544 Hepatitis
27	44	84.6	11	5	ABB80553	Abb80553 Hepatitis
28	44	84.6	11	5	ABB80552	Abb80552 Hepatitis
29	44	84.6	11	5	ABB80545	Abb80545 Hepatitis
30	42	80.8	11	5	ABB80530	Abb80530 Hepatitis
31	41	78.8	11	5	ABB80542	Abb80542 Hepatitis
32	41	78.8	11	5	ABB80543	Abb80543 Hepatitis
33	41	78.8	11	5	ABB80538	Abb80538 Hepatitis
34	40	76.9	11	5	ABB80548	Abb80548 Hepatitis
35	40	76.9	11	5	ABB80547	Abb80547 Hepatitis
36	40	76.9	11	5	ABB80556	Abb80556 Hepatitis
37	40	76.9	11	5	ABB80557	Abb80557 Hepatitis
38	40	76.9	11	5	ABB80537	Abb80537 Hepatitis
39	40	76.9	11	5	ABB80551	Abb80551 Hepatitis
40	40	76.9	11	5	ABB80541	Abb80541 Hepatitis
41	40	76.9	20	2	AAU76810	AAU76810 Hepatitis
42	40	76.9	1022	4	ABG03621	ABG03621 Novel hum
43	40	76.9	1022	4	ABG08173	ABG08173 Novel hum
44	40	76.9	1022	4	ABG05826	ABG05826 Novel hum
45	39	75.0	11	5	ABB80546	Abb80546 Hepatitis

ALIGNMENTS

RESULT 1

ABB80525

ID ABB80525 standard; peptide; 11 AA.

XX ABB80525;

AC ABB80525;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX virucide.

XX Synthetic.

XX Key

FH Location/Qualifiers

FT Modified-site 1

FT Modified-site /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

WO200208251-A2.

31-JAN-2002.

PF 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

PR (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

PI WPI; 2002-361643/39.

DR Novel peptide compound having hepatitis C virus protease inhibitory

XX activity useful for treating disorders associated with hepatitis C virus

PT protease.

XX Claim 17; Page 64; 69pp; English.

PS The sequence represents a peptide compound of the invention having

XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
 |||||
 DB 1 EVVVPXGMSYS 11

RESULT 2
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.
 AC ABB80521;
 XX 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
 XX Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.
 31-JAN-2002.
 19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.
 (CORV-) CORVAS INT INC.
 PI Lim-Wilby M, Levy OE, Brunck TK;
 WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.
 Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus
 SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EVVVPXGMSYS 11
 |||||
 DB 1 EVVVPXGMSYS 11

RESULT 3
 ABB80522
 ID ABB80522 standard; peptide; 11 AA.
 AC ABB80522;
 XX 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.

XX Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.

OS
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.
 31-JAN-2002.
 19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.
 (CORV-) CORVAS INT INC.
 PI Lim-Wilby M, Levy OE, Brunck TK;
 WPI; 2002-361643/39.
 Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

Claim 17; Page 64; 69pp; English.
 The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

SQ Sequence 11 AA;
 Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EVVVPXGMSYS 11
 |||||
 DB 1 EVVVPXGMSYS 11

RESULT 4

FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX Sequence 11 AA;
 SQ

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred.No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEWVPXGMSYS 11
 DB 1 EEWVPXGMSYS 11

RESULT 7
 ID ABB80567
 AC ABB80567 standard; peptide; 11 AA.
 AC ABB80567;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "(s/s)allothreonyl carbonyl residue forming a keto
 FT -amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX Sequence 11 AA;
 SQ

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred.No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEWVPXGMSYS 11
 DB 1 EEWVPXGMSYS 11

RESULT 8
 ID ABB80559
 AC ABB80559 standard; peptide; 11 AA.
 AC ABB80559;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
PS Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 9
ID ABB80526 standard; peptide; 11 AA.
AC ABB80526;
XX
DT 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PS Claim 17; Page 64; 69pp; English.
XX
SQ

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 10
ID ABB80564 standard; peptide; 11 AA.
AC ABB80564;
XX
DT 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PS Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
PS Claim 17; Page 65; 69pp; English.
XX
SQ The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 11
 ABB80568
 ID ABB80568 standard; peptide; 11 AA.
 XX AC ABB80568;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
 FT a keto-amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX PN 31-JAN-2002.
 XX PD 19-JUL-2001; 2001WO-US023169.
 XX PF 21-JUL-2000; 2000US-0220101P.
 XX PR (CORV-) CORVAS INT INC.
 XX PA Lim-Wilby M, Levy OE, Brunck TK;
 XX PI WPI; 2002-361643/39.
 XX DR Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 12
 ABB80561
 ID ABB80561 standard; peptide; 11 AA.
 XX AC ABB80561;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX PN 31-JAN-2002.
 XX PD 19-JUL-2001; 2001WO-US023169.
 XX PF 21-JUL-2000; 2000US-0220101P.
 XX PR (CORV-) CORVAS INT INC.
 XX PA Lim-Wilby M, Levy OE, Brunck TK;
 XX PI WPI; 2002-361643/39.
 XX DR Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 |||||
 Db 1 BEVVPXGMSYS 11

RESULT 13
 ABB80524
 ID ABB80524 standard; peptide; 11 AA.
 XX AC ABB80524;
 XX XX

DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
 DE
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 FT 31-JAN-2002.
 FT 19-JUL-2001; 2001WO-US023169.
 FT 21-JUL-2000; 2000US-0220101P.
 FT (CORV-) CORVAS INT INC.
 FT Lim-Wilby M, Levy OE, Brunck TK;
 FT WPI; 2002-361643/39.
 FT Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 FT Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0;
 QY 1 EEVVPXGMSYS 11
 |||||
 DB 1 EEVVPXGMDYS 11
 |||||
 RESULT 14
 ABB80529
 ID ABB80529 standard; peptide; 11 AA.
 XX
 AC ABB80529;
 XX
 DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
 DE
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 FT 31-JAN-2002.
 FT 19-JUL-2001; 2001WO-US023169.
 FT 21-JUL-2000; 2000US-0220101P.
 FT (CORV-) CORVAS INT INC.
 FT Lim-Wilby M, Levy OE, Brunck TK;
 FT WPI; 2002-361643/39.
 FT Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 FT Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013; Mismatches 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0;
 QY 1 EEVVPXGMSYS 11
 |||||
 DB 1 EEVVPXGMDYS 11
 |||||
 RESULT 15
 ABB80528
 ID ABB80528 standard; peptide; 11 AA.
 XX
 AC ABB80528;
 XX
 DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
 DE
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"

FT Misc-difference 8 residue 7"
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 XX 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 XX
 XX (CORV-) CORVAS INT INC.
 XX
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX
 XX Claim 17; Page 64; 69pp; English.
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EEVVPXGMSYS 11
 Db 1 EEVVPXGMDYS 11

Search completed: June 3, 2004, 11:48:23
 Job time : 45.9333 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	73.1	3472	4	US-09-408-020-4	Sequence 4, Appli
2	36	69.2	382	4	US-09-138-000C-3738	Sequence 3738, Ap
3	36	65.4	947	4	US-09-220-986-73	Sequence 73, Appl
4	34	65.4	1191	4	US-09-340-236-2902	Sequence 2902, Ap
5	34	65.4	1407	4	US-09-328-352-7885	Sequence 7885, Ap
6	33	63.5	12	4	US-09-760-946-2	Sequence 2, Appli
7	33	63.5	12	4	US-09-760-946-3	Sequence 3, Appli
8	33	63.5	45	2	US-08-637-759B-236	Sequence 236, App
9	33	63.5	45	3	US-08-871-355A-236	Sequence 236, App
10	33	63.5	45	4	US-09-201-945-236	Sequence 236, App
11	33	63.5	65	6	5177197-51	Patent No. 5177197
12	33	63.5	410	6	5177197-1	Patent No. 5177197
13	33	63.5	1394	6	5177197-30	Patent No. 5177197
14	32	61.5	10	3	US-09-357-952-66	Sequence 66, Appl
15	32	61.5	10	4	US-09-521-650-66	Sequence 66, Appl
16	32	61.5	10	4	US-09-168-888-66	Sequence 66, Appl
17	32	61.5	102	2	US-08-580-988A-23	Sequence 23, Appl
18	32	61.5	152	2	US-08-460-694-4	Sequence 4, Appli
19	32	61.5	152	3	US-08-460-744-4	Sequence 4, Appli
20	32	61.5	152	3	US-07-667-711B-4	Sequence 4, Appli
21	32	61.5	173	1	US-08-193-977-7	Sequence 7, Appli
22	32	61.5	189	2	US-08-464-517-21	Sequence 21, Appl
23	32	61.5	189	3	US-08-246-361A-21	Sequence 21, Appl
24	32	61.5	189	3	US-08-463-772-21	Sequence 21, Appl
25	32	61.5	189	5	PCT-US93-05000-21	Sequence 21, Appl
26	32	61.5	236	2	US-08-464-517-22	Sequence 22, Appl
27	32	61.5	236	2	US-08-246-361A-22	Sequence 22, Appl

FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (327)..(328)
OTHER INFORMATION: Amino acids 327 & 328 are Xaa wherein Xaa = any amino acid.
US-09-134-000C-3738

Query Match 69.2%; Score 36; DB 4; Length 382;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
: : : : :
Db 332 LIPEGMSYS 340

RESULT 3
US-09-228-986-73
Sequence 73, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-09-228-986-73

Query Match 65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 56.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
: : : : :
Db 686 VMPGMSYS 694

RESULT 4
US-09-540-236-2902
Sequence 2902, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2902
LENGTH: 1191
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2902

Query Match 65.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGMSYS 10
: : : : :
Db 783 EILPVGMAV 791

RESULT 5
US-09-328-352-7885

Sequence 7885, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7885
LENGTH: 1407
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match 65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGMSYS 10
: : : : :
Db 596 EVVPEGLSF 604

RESULT 6
US-09-760-946-2
Sequence 2, Application US/09760946
Patent No. 6608027
GENERAL INFORMATION:
APPLICANT: Teantrizos, Youla S.
APPLICANT: Cameron, Dale R.
APPLICANT: Faucher, Anne-Marie
APPLICANT: Ghio, Elise
APPLICANT: Goudreau, Nathalie
APPLICANT: Halmos, Teddy
APPLICANT: Llinas-Brunet, Montse
TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
FILE REFERENCE: 13/076-1-C1
CURRENT APPLICATION NUMBER: US/09/760,946
CURRENT FILING DATE: 2001-08-23
PRIOR APPLICATION NUMBER: US 09/542,675
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: US 60/128,011
PRIOR FILING DATE: 1999-04-06
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Substrate for recombinant HCV NS3 protease radiometric assay
US-09-760-946-2

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPGMSYS 11
: : : : :
Db 1 DDIVPCMSYS 11

RESULT 7
US-09-760-946-3
Sequence 3, Application US/09760946
Patent No. 6608027
GENERAL INFORMATION:
APPLICANT: Teantrizos, Youla S.
APPLICANT: Cameron, Dale R.
APPLICANT: Faucher, Anne-Marie
APPLICANT: Ghio, Elise

APPLICANT: Goudreau, Nathalie
APPLICANT: Halmos, Teddy
APPLICANT: Llinas-Brunet, Montse
TITLE OF INVENTION: Macrocytic Peptides Active Against the Hepatitis C Virus
FILE REFERENCE: 13/076-1-C1
CURRENT APPLICATION NUMBER: US/09/760,946
PRIOR FILING DATE: 2001-08-23
PRIOR APPLICATION NUMBER: US 09/542,675
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: US 60/128,011
PRIOR FILING DATE: 1998-04-06
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Tracer for NS3 protease assay
NAME/KEY: MOD_RES
LOCATION: (1)
OTHER INFORMATION: Asp at position 1 is biotinylated
NAME/KEY: MOD_RES
LOCATION: (10)
OTHER INFORMATION: Tyr at position 10 is iodinated with I-125
US-09-760-946-3

Query Match 63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 DRIVPCSMYSY 11

RESULT 8
US-08-637-759B-236
Sequence 236, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:

SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1 BEISPLGWSY 10

RESULT 9
US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 1 BEISPLGWSY 10

```
RESULT 10
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGWSY 10
Db 1 BEISPLGWSY 10

RESULT 11
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:51:
; LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGWSY 10
Db 1 BEISPLGWSY 10

RESULT 12
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:1:
; LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 97;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGWSY 11
Db 399 KEICPGMGYT 409

RESULT 13
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:30:
; LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGWSY 11
Db 399 KEICPGMGYT 409

RESULT 14
US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whole
; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Protease
; TITLE OF INVENTION: Other Enzymes and the Use Thereof
; FILE REFERENCE: 1735.0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642
```

; EARLIER FILING DATE: 21-JUL-1998
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
JS-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3; 2; Indels 0; Gaps 0;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

2y 1 BEVVPXGMSY 10
:::| | | |
db 1 DDIVPCMSY 10

RESULT 15
JS-09-521-650-66
; Sequence 66, Application US/09521650
; Patent No. 6335429
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.
; APPLICANT: Zhang, Han-Zhong
; TITLE OF INVENTION: No. 6335429a1 Fluorogenic or Fluorescent Reporter Molecules and
; TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
; TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1735.0290002
; CURRENT APPLICATION NUMBER: US/09/521,650
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 09/168,888
; EARLIER FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 60/061,582
; EARLIER FILING DATE: 1997-10-10
; EARLIER APPLICATION NUMBER: US 09/033,661
; EARLIER FILING DATE: 1998-03-03
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
:::| | | |
Db 1 DDIVPCMSY 10

Search completed: June 3, 2004, 12:03:07
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	12	US-09-909-164-5
2	50	96.2	11	12	US-09-909-164-6
3	50	96.2	11	12	US-09-909-164-9
4	50	96.2	11	12	US-09-909-164-10
5	50	96.2	11	12	US-09-909-164-47
6	50	96.2	11	12	US-09-909-164-48
7	50	96.2	11	12	US-09-909-164-49
8	50	96.2	11	12	US-09-909-164-50
9	50	96.2	11	12	US-09-909-164-51
10	50	96.2	11	12	US-09-909-164-52
11	46	88.5	11	12	US-09-909-164-8
12	46	88.5	11	12	US-09-909-164-12
13	46	88.5	11	12	US-09-909-164-13
14	45	86.5	11	12	US-09-909-164-7
15	45	86.5	11	12	US-09-909-164-11

16	45	86.5	11	12	US-09-909-164-19
17	45	86.5	11	12	US-09-909-164-20
18	45	86.5	11	12	US-09-909-164-23
19	45	86.5	11	12	US-09-909-164-24
20	44	84.6	11	12	US-09-909-164-28
21	44	84.6	11	12	US-09-909-164-29
22	44	84.6	11	12	US-09-909-164-33
23	44	84.6	11	12	US-09-909-164-36
24	44	84.6	11	12	US-09-909-164-37
25	44	84.6	11	12	US-09-909-164-43
26	42	80.8	11	12	US-09-909-164-14
27	41	78.8	11	12	US-09-909-164-22
28	41	78.8	11	12	US-09-909-164-26
29	41	78.8	11	12	US-09-909-164-27
30	41	78.8	11	12	US-09-909-164-61
31	41	78.8	11	12	US-09-909-164-62
32	40	76.9	11	12	US-09-909-164-21
33	40	76.9	11	12	US-09-909-164-25
34	40	76.9	11	12	US-09-909-164-31
35	40	76.9	11	12	US-09-909-164-32
36	40	76.9	11	12	US-09-909-164-35
37	40	76.9	11	12	US-09-909-164-41
38	40	76.9	11	12	US-09-909-164-41
39	40	76.9	11	12	US-09-909-164-45
40	39	75.0	11	12	US-09-909-164-46
41	39	75.0	11	12	US-09-909-164-30
42	39	75.0	11	12	US-09-909-164-34
43	39	75.0	11	12	US-09-909-164-38
44	39	75.0	11	12	US-09-909-164-39
45	39	75.0	11	12	US-09-909-164-42

ALIGNMENTS

RESULT 1

US-09-909-164-5
; Sequence 5, Application US/0909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURES:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-5

Query Match 96.2% Score 50; DB 12; Length 11;

```

; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: 09-909-164-9
;
; Query Match 96.2%; Score 50; DB 12; Length 11;
; Best Local Similarity 100.0%; Pred. No. 0.0014;
; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 1 EEVVPXGMSYS 11
| | | | | | | | | |
DB 1 EEVVPXGMSYS 11
;
RESULT 4
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US2002008702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunch, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEP
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)

```

Qy 1 EEVVPXGMSYS 11

Db 1 EVVVPXGMSYS 11
|||||

RESULT 8

US-09-909-164-50
; Sequence 50, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: 2-amino-butyric acid- (CO)
US-09-909-164-50

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 9

US-09-909-164-51
; Sequence 51, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 10

US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 11

US-09-909-164-8
; Sequence 8, Application US/09909164
; Publication No. US20020068702A1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: (s,s)-allothreonine-(CO)
US-09-909-164-51

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 10

US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 11

US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
|||||

Db 1 EVVVPXGMSYS 11
|||||

RESULT 14

```
US-09-909-164-7
; Sequence 7, Application US/0909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-7
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 BEVVPXGMSYS 11
   |||||
Db 1 BEVVPXGMHYS 11
```

RESULT 15

```
US-09-909-164-11
; Sequence 11, Application US/0909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
```

```
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (2)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-11
```

```
Query Match      86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 BEVVPXGMSYS 11
   |||||
Db 1 BEVVPXGMHYS 11
```

```
Search completed: June 3, 2004, 12:57:15
Job time : 33.7333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

run on: June 3, 2004, 11:35:47 ; Search time 9 seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2 T31308	hypothetical 367K
2	37	71.2	840	2 T39116	probable sulfate p
3	37	71.2	877	2 T40413	sulfate permease -
4	36	69.2	102	2 A42452	V1 protein - tobac
5	36	69.2	1498	2 B37355	DNA segregation AT
6	35	67.3	225	2 S57810	hypothetical prote
7	35	67.3	425	2 T24111	hypothetical prote
8	35	67.3	670	2 S22293	zinc finger protei
9	35	67.3	749	2 H82691	topoisomerase IV s
10	35	67.3	2717	2 A34203	DNA-binding protei
11	34	65.4	156	2 S54619	hypothetical prote
12	34	65.4	252	2 H69491	cell division inh
13	34	65.4	544	2 C82900	probable ABC subst
14	33	63.5	94	2 I40758	hypothetical prote
15	33	63.5	116	2 E20544	50S ribosomal prot
16	33	63.5	165	2 D59493	hypothetical prote
17	33	63.5	253	2 C81374	hypothetical prote
18	33	63.5	259	2 T34536	hypothetical prote
19	33	63.5	284	2 S75817	hypothetical prote
20	33	63.5	298	2 T47670	beta-ketoacyl-ACP
21	33	63.5	368	2 F72281	hypothetical prote
22	33	63.5	426	2 D82163	3-phosphoshikimate
23	33	63.5	466	2 T34653	cdc37 protein - fi
24	33	63.5	653	2 D82352	iron(III) ABC tran
25	33	63.5	890	2 A30481	bacteriocin BCN5 -
26	33	63.5	1028	2 AF3286	ATP-dependent DNA
27	33	63.5	1152	2 D87046	conserved hypothet
28	33	63.5	1194	2 A35626	transforming growt
29	33	63.5	1401	2 G82336	DNA-directed RNA p

ALIGNMENTS

RESULT 1

T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C/Species: Cenarchaeum symbiosum
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C/Accession: T31308
R/Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A/Title: Genomic analysis reveals chromosomal variation in natural populations of the un
A/Reference number: Z20994; MUID:98422450; PMID:9748430
A/Accession: T31308
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3472 <SCH>
A/Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC62699.1
C/Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. No. 60;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11

Db 2294 EDVIPRGISFS 2304

RESULT 2

T39116
probable sulfate permease - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T39116
R/Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A/Reference number: Z21829
A/Accession: T39116
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-840 <HUN>
A/Cross-references: EMBL:AL132779; PIDN:CA860015.1; GSPDB:GN00066; SPDB:SPAC869.05c
A/Experimental source: strain 972h-; cosmid c869
C/Genetics:
A/Gene: SPDB:SPAC869.05c
A/Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

Db 135 VVPQMSYA 143

hypothetical prote
masking protein pr
hypothetical prote
transport protein
rho protein GDP-di
hypothetical prote
cyclin D2 - rat
cyclin D2 - mouse
cyclin D2 - human
cyclin D1 - Africa
cyclin D1 - Africa
cyclin D2 - chicke
cyclin D1 - zebra
cyclin D3 - human
cyclin D1 - human

RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-877 <LYN>
A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A:Experimental source: strain 972h-; cosmid c3H7
C:Genetics:
A:Gene: SPDB:SPBC3H7.02
A:Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
||| |||
Db 148 VVFGMSYA 156

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <VOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
: ||| : |||
Db 7 QVFPNGINS 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:g15026814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
: : |||
Db 1276 EQIPMGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TP11) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of pistil-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MIL>
A:Cross-references: EMBL:U20592; NID:g924625; PIDN:AAA80497.1; PID:g924625
C:Superfamily: plant kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
: ||| : |||
Db 32 DEVVPGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WIL>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
: ||| : |||
Db 335 EQIVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; 178656
R:Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

A;Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
A;Reference number: 158280; MUID:91187610; PMID:1901405
A;Accession: S22293
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-670 <MT>
A;Cross-references: EMBL:X54250; NID:g57519; PIDN:CAA38151.1; PID:g57520
A;Note: the authors did not translate the codon for residue 1
A;Superfamily: HIV-EP2 enhancer-binding protein
A;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DY 3 VVPXGMSYS 11
DB 376 VVPAGLYTS 384
|||||
|||||

RESULT 9
82691
:topoisomerase IV subunit XF1353 [imported] - Xylella fastidiosa (strain 9a5c)
A;Species: Xylella fastidiosa
A;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
A;Accession: H82691
A;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82691
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-749 <STM>
A;Cross-references: GB:AB003967; GB:AB003849; NID:g9106347; PIDN:AAF84162.1; GSPDB:GN001
A;Experimental source: strain 9a5c
A;Simpson: A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-
briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrex, H-
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm-
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig-
shado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E-
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A-
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.Z.; de Sa, R.G.; Santelli, R.V.; Sawasak-
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir-
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z-
A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF1353
A;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase C

Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DY 2 EVVPXGMSY 10
DB 526 EVDPGMSY 534
|||||
|||||

RESULT 10
A34203
DNA-binding protein PRDIT-BF1 - human
A;Alternate names: major histocompatibility complex enhancer-binding protein 1
A;Species: Homo sapiens (man)
A;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
A;Accession: A34203; A34779
A;Pan, C.M.; Maniatis, T.
Genes Dev. 4, 29-42, 1990

A;Title: A DNA-binding protein containing two widely separated zinc finger motifs that i-
A;Reference number: A34203; MUID:90169514; PMID:2106471
A;Accession: A34203
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2717 <FAN>
A;Cross-references: EMBL:X51435; NID:g38017; PIDN:CAA35798.1; PID:g38018
A;Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
A;Title: A large protein containing zinc finger domains binds to related sequence elemen-
A;Reference number: A34779; MUID:90205817; PMID:2108316
A;Accession: A34779
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-16
A;Cross-references: GB:M32019
A;Superfamily: HIV-EP2 enhancer-binding protein
A;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DY 3 VVPXGMSYS 11
DB 2405 VVPAGLYTS 2413
|||||
|||||

RESULT 11
S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
A;Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
A;Species: Saccharomyces cerevisiae
A;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
A;Accession: S54619; S66879
A;Cross-references: EMBL:X87331; NID:g1041552; PIDN:CAA60762.1; PID:g829123
A;Rde Haan, M.; Maarse, A.C.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66877
A;Accession: S66879
A;Molecule type: DNA
A;Residues: 1-156 <DBH>
A;Cross-references: EMBL:X87331; NID:g1041552; PIDN:CAA60762.1; PID:g829123
A;Rde Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66877
A;Accession: S66879
A;Molecule type: DNA
A;Residues: 1-156 <DBH>
A;Cross-references: EMBL:X87331; NID:g1041552; PIDN:CAA60762.1; PID:g829123
A;Experimental source: strain S288C
A;Genetics:
A;Cross-references: SGD:S0005539
A;Map position: 15R
A;Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DY 2 EVVPXGMSY 10
DB 50 EVVPLGMDY 58
|||||
|||||

RESULT 12
H69491
cell division inhibitor (mind-2) homolog - Archaeoglobus fulgidus
A;Species: Archaeoglobus fulgidus
A;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
A;Accession: H69491
A;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Wosee, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AE000970; GB:AE000782; NID:G2689293; PIDN:AA89318.1; PID:G264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMS 9
 Db 81 EVIPAGMS 88

RESULT 13

C82900
 Probable ABC substrate-binding protein, iron U359 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:G6899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCsbp-5; U359
 A:Genetic code: GCG

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVFXGMSY 10
 Db 135 EVVPHYLSY 144

RESULT 14

I40758
 Hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohyd
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:Z36940; NID:G535805; PIDN:CAA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 10
 Db 26 DIFPSGMSY 34

RESULT 15

E90544
 50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Sanson, D.; Gallison, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma puli
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: E90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPU 2610
 A:Genetic code: GCG
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
 Db 68 VVFLGMSYS 76

Search completed: June 3, 2004, 12:00:00
 Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 3, 2004, 11:32:06 ; Search time 4.86667 seconds
(without alignments)
117.693 Million cell updates/sec

title: US-09-909-164-10
perfect score: 52
sequence: 1 EEVVPXGMSYS 11

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	1058	1 CARB_FUSNN	Q8rg66 fusobacteri
2	37	71.2	877	1 SULH_SCHPO	O74377 schizosacch
3	36	69.2	102	1 V1LK_TYDVA	P31619 tobacco yel
4	36	69.2	1498	1 V1A9_CLOAB	Q04351 clostridium
5	35	67.3	2717	1 ZEP1_HUNAN	P15822 homo sapien
6	34	65.4	788	1 CY14_NEUCR	P23822 neurospora
7	34	65.4	1499	1 A10A_HUMAN	O60312 homo sapien
8	33	63.5	116	1 RL20_MYCFU	Q98q70 mycoplasma
9	33	63.5	165	1 V149_ARCFU	O28330 archaeoglob
10	33	63.5	253	1 V990_CMJTE	P45489 campylobact
11	33	63.5	280	1 CUX3_MOUSE	Q9d387 mus musculu
12	33	63.5	426	1 AKO4_VIBCH	Q9k2b0 vibrio chol
13	33	63.5	466	1 CC37_SCHPO	O94740 schizosacch
14	33	63.5	478	1 GSR2_HUMAN	Q9n2m5 homo sapien
15	33	63.5	890	1 BGN5_CLOPE	P08696 clostridium
16	33	63.5	1389	1 LTB5_MOUSE	O8c918 mus musculu
17	33	63.5	1394	1 LTB5_MOUSE	P22064 homo sapien
18	33	63.5	1401	1 RPO3_VIBCH	Q9kv29 vibrio chol
19	33	63.5	1595	1 LTB1_HUMAN	Q14766 homo sapien
20	33	63.5	1712	1 LTB1_RAT	Q00818 rattus norv
21	33	63.5	1713	1 LTB1_MOUSE	Q8c919 mus musculu
22	32	61.5	288	1 CGD2_RAT	Q04827 rattus norv
23	32	61.5	289	1 CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	1 CGD2_MOUSE	P30280 mus musculu
25	32	61.5	291	1 CGD1_BRARE	P50755 xenopus lae
26	32	61.5	291	1 CGD1_XENLA	P49706 gallus gall
27	32	61.5	291	1 CGD2_CHICK	P53782 xenopus lae
28	32	61.5	291	1 CGD2_XENLA	P55169 gallus gall
29	32	61.5	292	1 CGD1_CHICK	P30281 homo sapien
30	32	61.5	292	1 CGD3_HUMAN	P24385 homo sapien
31	32	61.5	295	1 CGD1_HUMAN	P25322 mus musculu
32	32	61.5	295	1 CGD1_MOUSE	P39948 rattus norv
33	32	61.5	295	1 CGD1_RAT	

RESULT 1
CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG66;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatzal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardier W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A., Larsen N., D'Souza M., Malunas T., Fusch G., Haselkorn R., Ponstein M., Kyrpides N., Overbeek R., RA
RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586." J. Bacteriol. 184:2005-2018(2002).
RL -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -1- SIMILARITY: Belongs to the carb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AEO10554; AAL94625.1; ALT_INTT.
CC HAMAP; MF_01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_Like.
DR Pfam; PF00289; CPsase_L_chain; 2.
DR Pfam; PF02786; CPsase_L_D2; 2.
DR Pfam; PF02787; CPsase_L_D3; 1.

P44677 haemophilus
O75355 homo sapien
O9y616 homo sapien
P90518 crithidia f
P52384 human herpe
P52544 human herpe
P32784 saccharomyc
Q62671 rattus norv
P28931 tomato aspe
P16916 escherichia
P16918 escherichia
P16917 escherichia

ALIGNMENTS

pfam; PF02142; MGS; 1.
 PRINTS; PRO0098; CPSASE.
 TIGRFAMs; TIGR01369; CPSase1_lrg; 1.
 PROSITE; PS00866; CPSASE 1; 2.
 PROSITE; PS00867; CPSASE 2; 2.
 Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401
 FT DOMAIN 402 546
 FT DOMAIN 547 929
 FT DOMAIN 930 1058
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210
 FT NP_BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;
 Query Match 73.1%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 7;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 EVVPGXGMSYS 11
 Db 190 EIVPGNGLYS 199
 |||||:|
 RESULT 2
 SULH SCHPO STANDARD; PRT; 877 AA.
 AC 074377;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable sulfate permease CJH7.02.
 GN SPBC3H7.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OK NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwalliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis F., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle B.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Welljens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Drano S., Gloux S., Delaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415;871-880(2002).
 CC -!- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the SLC26A/SulP transporter (TC 2.A.53)
 CC family.
 CC -!- SIMILARITY: Contains 1 STAS domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL031261; CA20298.1; -.
 CC FIR; T40413; T40413.
 CC GenBank; SPombe; SPBC3H7.02; -.
 CC InterPro; IPR002645; STAS.
 CC InterPro; IPR001902; Sulph_transpt.
 CC Pfam; PF01740; STAS; 1.
 CC Pfam; PF00916; Sulfate_transp; 1.
 CC TIGRFAMs; TIGR00815; sulP; 1.
 CC PROSITE; PS01130; SLC26A; 1.
 CC PROSITE; PS0801; STAS; 1.
 CC Transport; Transmembrane.
 CC TRANSMEM 133 153 POTENTIAL.
 CC TRANSMEM 161 181 POTENTIAL.
 CC TRANSMEM 186 206 POTENTIAL.
 CC TRANSMEM 221 241 POTENTIAL.
 CC TRANSMEM 243 263 POTENTIAL.
 CC TRANSMEM 292 312 POTENTIAL.
 CC TRANSMEM 329 349 POTENTIAL.
 CC TRANSMEM 384 404 POTENTIAL.
 CC TRANSMEM 424 444 POTENTIAL.
 CC TRANSMEM 461 481 POTENTIAL.
 CC TRANSMEM 484 504 POTENTIAL.
 CC TRANSMEM 518 538 POTENTIAL.
 CC TRANSMEM 543 563 POTENTIAL.
 CC TRANSMEM 594 747 STAS.
 CC DOMAIN 594 747 STAS.
 CC SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;
 SQ
 Query Match 71.2%; Score 37; DB 1; Length 877;
 Best Local Similarity 77.8%; Pred. No. 9.4;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMSYS 11
 Db 148 VVPQGNYSVA 156
 |||||:
 RESULT 3
 Y1LK TYDVA STANDARD; PRT; 102 AA.
 ID Y1LK TYDVA
 AC P1619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 DE V.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 GN Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 OC NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92188538; PubMed=1546458;
 RX Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RA "The nucleotide sequence of the infectious cloned DNA component of
 RA tobacco yellow dwarf virus reveals features of geminiviruses
 RA infecting monocotyledonous plants.";
 RT Virology 187:633-642(1992).
 RL

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; M81103; AAA47947.1; --
 PIR; A42452; A42452.
 InterPro; IPR002621; Gemini.mov.
 Pfam; PF01708; Gemini.mov; 1.
 Hypothetical protein.

Query Match 69.2%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 1.6;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Y 2 EVVPXGMSYS 11
 b 7 QVPSGINS 16

RESULT 4
 YIA9_CLOAB STANDARD; PRT; 1498 AA.
 ID Q04351;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CAC3709.
 GN CAC3709.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 XC Clostridium.
 NCBI_TaxID=1488;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 MEDLINE=21359325; PubMed=11466286;
 Noelling J., Berton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
 Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 Bennett G.N., Koonin E.V., Smith D.R.;
 "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.";
 J. Bacteriol. 183:4823-4838(2001).
 [2]
 SEQUENCE OF 1-108 FROM N.A.
 STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 MEDLINE=93273706; PubMed=8501044;
 Sauer U., Duerre P.;
 "Sequence and molecular characterization of a DNA region encoding a small heat shock protein of Clostridium acetobutylicum.";
 J. Bacteriol. 175:3394-3400(1993).
 -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts in positions 76 and 106.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AE007866; AA891629.1; --
 PIR; X65276; CAA46379.1; ALT FRAME.
 InterPro; IPR002543; FtsK_SpoIIIE.

Pfam; PF01580; FtsK_SpoIIIE; 2.
 DR PROSITE; PS50901; FTSK; 2.
 KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
 FT DOMAIN 655 857 FTSK 1.
 FT DOMAIN 1001 1188 FTSK 2.
 FT NP_BIND 675 682 ATP (POTENTIAL).
 SQ SEQUENCE 1498 AA; 168968 MW; FF42037A335A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
 Best Local Similarity 60.0%; Pred. No. 27;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Q 1 EVVPXGMSY 10
 b 1276 EQKIPMGMSY 1285

RESULT 5
 ZEPI_HUMAN STANDARD; PRT; 2717 AA.
 ID ZEPI_HUMAN
 AC P15822;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-binding protein 1) (HIV-EP1) (Major histocompatibility complex binding protein 1) (MBP-1) (Positive regulatory domain II binding factor 1) (PRDII-BF1).
 GN HIVEP1 OR ZNF40.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=90169514; PubMed=2106471;
 Fan C.M., Maniatis T.;
 "A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence.";
 Genes Dev. 4:29-42(1990).
 [2]
 STRUCTURE BY NMR OF 2113-2142.
 MEDLINE=91084333; PubMed=2248949;
 Omichinski J.G., Clore G.M., Appella E., Sakaguchi K., Gronenborn A.M.;
 "High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution.";
 Biochemistry 29:9324-9334(1990).
 [3]
 STRUCTURE BY NMR OF 2087-2142.
 MEDLINE=92232684; PubMed=1567844;
 Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E., Gronenborn A.M.;
 "High-resolution solution structure of the double CysHis2 zinc finger from the human enhancer binding protein MBP-1.";
 Biochemistry 31:3907-3917(1992).
 -!- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE 5'-GGGATTTC-3', WHICH IS FOUND IN THE ENHANCER ELEMENTS OF NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1. IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT IN T-CELL ACTIVATION.
 -!- SUBCELLULAR LOCATION: Nuclear.
 -!- INDUCTION: By mitogens and phorbol ester.
 -!- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.
 -!- SIMILARITY: STRONG, TO HIVP2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X51435; CAA35798.1; --
 PIR; A34203; A34203.
 PDB; 3ZF; 15-JAN-92.
 PDB; 4ZNF; 15-JAN-92.
 PDB; 1BBO; 31-OCT-93.
 TRANSFAC; T00497; --
 GenBank; HGNC:4920; HIVEP1.
 MIM; 194540; --
 GO; GO:0005634; C:nucleus; TAS.
 GO; GO:0003677; F:DNA binding; TAS.
 InterPro; IPR007087; Znf_C2H2.
 Pfam; PF00096; zf-C2H2; 5.
 SMART; SM00355; Znf_C2H2; 4.
 PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
 Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 Nuclear protein; Repeat; 3D-structure.
 ZN_FING 406 428
 ZN_FING 434 456
 ZN_FING 434 456
 ZN_FING 958 981
 ZN_FING 2087 2109
 ZN_FING 2115 2139
 ZN_FING 803 806
 DOMAIN 2088 2088
 STRAND 2090 2092
 TURN 2090 2092
 STRAND 2095 2095
 HELIX 2099 2108
 TURN 2109 2109
 STRAND 2115 2116
 STRAND 2123 2124
 HELIX 2127 2135
 SEQUENCE 2177 AA; 297217 MW; D45D3CA951FPA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
 Db 2405 VVPAGLTYS 2413

RESULT 6
 CY14_NEUCR
 ID CY14_NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate permease II.
 CYS-14.
 GN Neurospora crassa.
 OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 NCBI_TaxID=5141;
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=91129256; PubMed=1825178;
 RA Ketter J.S., Jarai G., Fu Y.-H., Marzluf G.A.; and DNA recognition
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of *cys-14*, the structural gene for sulfate permease II in
 RT Neurospora crassa."
 RL Biochemistry 30:1780-1787(1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RA MEDLINE=94186926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate

permease II and a putative human tumour suppressor.";
 Trends Biochem. Sci. 19:19-19(1994).
 CC -!- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- INDUCTION: Highly expressed, but only in cells subject to sulfur
 CC limitation, and it is turned on by the positive-acting Cys-3
 CC sulfur regulatory protein.
 CC -!- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC -!- SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53)
 CC family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

EMBL; M59167; AAA33615.1; ALT_SEQ.
 PIR; A37956; A37956.
 InterPro; IPR001902; Sulph. transpt.
 Pfam; PF00916; Sulfate transp; 1.
 TIGRFAMs; TIGR00815; sulp; 1.
 PROSITE; PS01130; SLC26A; 1.
 Transport; Transmembrane; Glycoprotein.
 TRANSMEM 71 91
 TRANSMEM 103 123
 TRANSMEM 128 148
 TRANSMEM 171 191
 TRANSMEM 193 213
 TRANSMEM 271 291
 TRANSMEM 326 346
 TRANSMEM 363 383
 TRANSMEM 451 471
 TRANSMEM 474 494
 CARBOHYD 23 23
 CARBOHYD 578 578
 SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
 Db 90 VVPQWYAY 98

RESULT 7
 A10A_HUMAN
 ID A10A_HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; Q96914.
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
 DE (Aminophospholipid translocase VA).
 GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21225279; PubMed=11326269;
 RX Meguro M., Kashiwagi A., Mitsuya K., Nakao M., Kondo I., Saitoh S.,
 RA Osimura M.;
 RA "A novel maternally expressed gene, ATP10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome.";
 RL Nat. Genet. 28:19-20(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

DR	EMBL; AY029493; AAK33100.1; JOINED.
DR	EMBL; AY029494; AAK33100.1; JOINED.
DR	EMBL; AY029495; AAK33100.1; JOINED.
DR	EMBL; AY029496; AAK33100.1; JOINED.
DR	EMBL; AY029497; AAK33100.1; JOINED.
DR	EMBL; AY029498; AAK33100.1; JOINED.
DR	EMBL; AY029499; AAK33100.1; JOINED.
DR	EMBL; AY029500; AAK33100.1; JOINED.
DR	EMBL; AY029501; AAK33100.1; JOINED.
DR	EMBL; AY029502; AAK33100.1; JOINED.
DR	EMBL; BC029503; AAK33100.1; --
DR	EMBL; BC052251; AAH52251.1; --
DR	EMBL; AB011138; BRA25492.1; --
DR	Genew; HGNC:13542; ATP10A.
DR	MIM; 605855; --
DR	MIM; 105830; --
DR	GO; GO:0016021; C:integral to membrane; NAS.
DR	GO; GO:0004012; F:phospholipid-translocating. ATPase activity; NAS.
DR	GO; GO:0008360; P:regulation of cell shape; NAS.
DR	InterPro; IPR001757; ATpase_E1-E2.
DR	InterPro; IPR06539; Flippase.
DR	InterPro; IPR005834; Hydrolase.
DR	Pfam; PF00702; Hydrolase; 1.
DR	PRINTS; PR00119; CATATPASE.
DR	TIGRFAMs; TIGR01652; ATpase_Plipid; 1.
DR	TIGRFAMs; TIGR01494; ATpase_P-type; 6.
DR	PROSITE; PS00154; ATpase_E1_E2; 1.
KW	Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW	Multicene family.
FT	DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 87 106 POTENTIAL.
FT	DOMAIN 107 110 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 111 128 POTENTIAL.
FT	DOMAIN 129 309 POTENTIAL.
FT	TRANSMEM 310 332 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 333 362 POTENTIAL.
FT	TRANSMEM 363 384 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 385 1087 POTENTIAL.
FT	TRANSMEM 1088 1108 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 1109 1119 POTENTIAL.
FT	TRANSMEM 1120 1140 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 1141 1170 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 1171 1192 POTENTIAL.
FT	DOMAIN 1193 1199 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1200 1222 POTENTIAL.
FT	DOMAIN 1223 1228 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 1229 1249 POTENTIAL.
FT	DOMAIN 1250 1267 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1268 1292 POTENTIAL.
FT	DOMAIN 1293 1499 CYTOPLASMIC (POTENTIAL).
FT	MOD RES 427 427 PHOSPHORYLATION (BY SIMILARITY).
FT	METAL 1031 1031 MAGNESIUM (BY SIMILARITY).
FT	METAL 1035 1035 MAGNESIUM (BY SIMILARITY).
FT	DOMAIN 467 470 POLY-GLU.
FT	CONFLICT 388 388 Q -> R (IN REF. 4).
FT	SEQUENCE 1499 AA; 167687 MW; D4996AA4D0635A68D CRC64;
SQ	
Query Match	65.4%; Score 34; DB 1; Length 1499;
Best Local Similarity	72.7%; Pred. No. 70;
Matches	8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy	1 EEVVPXGMSYS 11
Db	469 EEVVPRGGSVS 479
RESULT 8	
RL20 MYCPU	
ID_RL20_MYCPU	STANDARD;
AD AC	PRT; 116 AA.
IC Q980V0:	
DT 28-FEB-2003	(Rel. 41, Created)
DT 28-FEB-2003	(Rel. 41, Last sequence update)
DT 28-FEB-2003	(Rel. 41, Last annotation update)

DE 50S ribosomal protein L20.
 GN RLPT OR MYPU 2610.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 CX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Moszer I., Dybwig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis."
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50s ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: Belongs to the L20P family of ribosomal proteins.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL445563; CAC13434.1; -.
 DR PIR; E90544; E90544.
 DR MyPulList; MYPU 2610; -.
 DR HAMAP; MF 00382; -; 1.
 DR InterPro; IPR005813; Ribosomal L20.
 DR InterPro; IPR005812; Ribosomal_L20b/o.
 DR Pfam; PF00453; Ribosomal_L20; 1.
 DR PRINTS; PRO0062; RIBOSOMALL20.
 DR ProDom; PD002389; L20; 1.
 DR TIGRFAMs; TIGR01032; rplT_bact; 1.
 DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 SQ SEQUENCE 116 AA; 13565 MW; C59C748901B1814 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 116;
 Best Local Similarity 77.8%; Pred. No. 8; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 2;
 QY 3 VVPXGMSYS 11
 DB 68 VRPLGMSYS 76
 RESULT 9
 ID YU49_ARCFU STANDARD; PRT; 165 AA.
 AC 028330;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein AF1949.
 GN AF1949.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 CX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Klenk H.-P., Clayton R.A., Gwin M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus."
 RL Nature 390:364-370(1997).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000969; AAB89307.1; -.
 DR PIR; D69493; D69493.
 DR TIGR; AF1949; -.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 SQ SEQUENCE 165 AA; 17588 MW; B5C17054810ADB8 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 165;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEWPKGMSY 10
 DB 60 EBSIPDGASY 69
 RESULT 10
 ID Y990_CAMJE STANDARD; PRT; 253 AA.
 AC P45489; O9PNV0;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein Cj0990C.
 GN Cj0990C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 CX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 RN [2]
 RP SEQUENCE OF 160-253 FROM N.A.
 RC STRAIN=ATCC 33431 / TGH 9011;
 RX MEDLINE=95247673; PubMed=7730270;
 RA Hani E.K., Chan V.L.;
 RT "Expression and characterization of Campylobacter jejuni
 RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
 RT coli."
 RL J. Bacteriol. 177:2396-2402(1995).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AL139076; CAB73246.1; -
 DR EMBL; Z36940; CA85392.1; -
 DR FIR; C81374; C81374.
 DR FIR; I40758; I40758.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
 Best Local Similarity 55.6%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXGMSY 10
 Db 185 DFFPGMSY 193

RESULT 11
 ID CTX3 MOUSE STANDARD; PRT; 280 AA.
 AC Q9D387; Q9CXQ4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein C2orf103 homolog precursor.
 GN C200RF103.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordore P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [2]

SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max J., Wang J., Haie F.,
 RA Diatchenko L., Marcus H., Moore T., Max J., Wang J., Haie F.,
 RA Diatchenko L., Marcus H., Moore T., Max J., Wang J., Haie F.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshlyuk S., Carninci P., Franke C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Griewood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Kryzinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AK014127; BAB29169.1; -
 DR EMBL; AK018222; BAB31124.1; -; ALT_FRAME.
 DR EMBL; BC004791; AAH04791.1; -
 DR MGD; MGI:1920368; 3110035N03Rik.
 DR MGD; MGI:1923411; 6330527O06Rik.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 280 PROTEIN C200RF103 HOMOLOG.
 FT DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 236 256 POTENTIAL.
 FT DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 1 118 Missing (in isoform 2).
 FT CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).
 FT CONFLICT 230 230 Q -> P (IN REF. 1; BAB31124).
 FT CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).
 SQ SEQUENCE 280 AA; 31721 MW; FA1D7BF9FD5CCEFCRC64;

Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPGXGMSY 10
 Db 173 VTPAGMSY 180

RESULT 12
 ID AROA VIBCH STANDARD; PRT; 426 AA.
 AC Q9KRE0.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR VCI1732.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;

[1] SEQUENCE FROM N.A.
 RN STRAIN=El Tor N16961 / Serotype O1;
 RC MEDLINE=20406833; PubMed=10952201;
 RX Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bases S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RT Nature 406:477-483 (2000).
 RL -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE004251; AAP94882.1; --
 DR PIR; D82163; D82163.
 DR TIGR; VCI732; --
 DR HAMAP; MF 00210; -- 1.
 DR InterPro; IPR006264; AroA.
 DR InterPro; IPR001986; EPSP synth.
 DR Pfam; PF00275; EPSP synthase; 1.
 DR ProDom; PD001867; EPSP synthase; 1.
 DR TIGRFAMs; TIGR01356; aroA; 1.
 DR PROSITE; PS00104; EPSP SYNTHASE_1; 1.
 DR PROSITE; PS00885; EPSP SYNTHASE_2; 1.
 DR KQ Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 426 AA; 46101 MW; 38852D6483BF1C3 CRC64;

 Query Match 63.5%; Score 33; DB 1; Length 426;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EHVVPXGMSY 10
 DB 223 EFVIPAGQSY 232

 RESULT 13
 CC37_SCHPO STANDARD; PRT; 466 AA.
 ID CC37_SCHPO
 AC 094740;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
 DE subunit) (Cell division control protein 37).
 GN CDC37 OR SPBC9B6.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Westwood P.K., Preston N.C., Fantes P.A.;
 RT "Schizosaccharomycetes pombe cdc37 gene."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2] SEQUENCE FROM N.A.
 RN STRAIN=972;
 RC MEDLINE=21848401; PubMed=11859360;
 RX Wood V., Guilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman C.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Howarth T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Meule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Warren T., Whitehead S.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Grymonprez B.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Schaefer M., Mueller-Auer S.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Hilbert H.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
 RT "The genome sequence of *Schizosaccharomyces pombe*."
 CC -----
 CC Nature 415:871-880 (2002).
 RN [3]
 RN FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RP MEDLINE=22745456; PubMed=12861001;
 RX Tatebe H., Shiozaki K.;
 RA Identification of cdc37 as a novel regulator of the stress-responsive
 RT mitogen-activated protein kinase.
 RL Mol. Cell. Biol. 23:5132-5142 (2003).
 CC -1- FUNCTION: Co-chaperone that binds to numerous kinases and promotes
 CC their interaction with the Hsp90 complex, resulting in
 CC stabilization and promotion of their activity.
 CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with styl.
 CC -1- SUBCELLULAR LOCATION: Nuclear, and cytoplasmic. When in the
 CC nucleus associated with chromatin.
 CC -1- SIMILARITY: Belongs to the CDC37 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ132377; CAB38758.1; --
 DR EMBL; AJ132376; CAB38757.1; --
 DR EMBL; AL049769; CAB42371.2; --
 DR PIR; T43653; T43653
 DR GeneDB_Spombe; SPBC9B6.10; --
 DR InterPro; IPR004918; Cdc37.
 DR Pfam; PF03234; Cdc37; 1.
 KW Chaperone; Cell division; Cell cycle; Nuclear protein.
 SQ SEQUENCE 466 AA; 52554 MW; 647238B34CABB3C5 CRC64;

 Query Match 63.5%; Score 33; DB 1; Length 466;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EHVVPXGMSY 10
 DB 98 DSAIPGMSY 107

 RESULT 14

3SR2_HUMAN
ID GSR2_HUMAN STANDARD; PRT; 478 AA.
AC Q9NZM5, Q9BT66; Q9HAX6; Q9NP1; Q9NPR4; Q9UF12;
JT 16-OCT-2001 (Rel. 40, Created)
JT 16-OCT-2001 (Rel. 40, Last sequence update)
JT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
EN GLTSCR2.
DS Homo sapiens (Human).
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RA MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramasamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.,
RA "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RA region.";
RT Genomics 64:44-50 (2000).
RL [2]
RN SEQUENCE FROM N.A.
RP SEQUENCE-Muscle;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Bruni R., Fineschi B., Ogle W.O., Roizman B.,
RA "A novel cellular protein, p60, interacting with both herpes simplex
RA virus 1 regulatory proteins ICP22 and ICP0 is modified in a
RA cell-type-specific manner and is recruited to the nucleus after
RA infection.";
RT J. Virol. 73:3810-3817 (1999).
RL [4]
RN SEQUENCE OF 12-478 FROM N.A.
RP SEQUENCE OF 12-478 FROM N.A.
RA Andreu N., Estivill X., Escarceller M., Suncy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 218-477 FROM N.A.
RC TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed at high levels in heart and
CC pancreas, moderate levels in placenta, liver, skeletal muscle, and
CC kidney, and low levels in brain and lung.
CC -1- SIMILARITY: Belongs to the GLTSCR2 family.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF182076; AAF62873.1; -
DR EMBL; BC004229; AAH04229.1; -
DR EMBL; BC006311; AAH06311.1; -
DR EMBL; BC010095; AAH10095.1; -
DR EMBL; AF296124; AAG30413.1; -
DR EMBL; AL359335; CAB94786.1; -
DR EMBL; AL359336; CAB94787.1; -
DR EMBL; AL122063; CAB59242.1; -
DR SWISS-2DPAGE; Q9NZM5; HUMAN.
DR Genew; HGNC:4333; GLTSCR2.
DR MIN; 605691; -
DR GO; GO:0005622; C:intracellular; NAS.
DR Nuclear protein; Polymorphism.
KW VARIANT 389 389 R -> Q.
FT CONFLICT 4 6 /FTID=VAR 011486.
FT CONFLICT 9 9 GGS -> HEG (IN REF. 2; AAH04229).
FT CONFLICT 146 191 G -> R (IN REF. 3).
FT CONFLICT 198 215 RRKEQLWEKLAKQGLPREVRAQARLLNPSATRAKPGFQD
FT CONFLICT 235 235 TVERP -> SGRSSYGRSWPSRASSPGGAQSPVAQPCN
FT CONFLICT 417 417 XGENPAPGRHIAA (IN REF. 3).
FT CONFLICT 433 477 SUNPLDRPLVQDEFFLE -> LNNPDKPVVWPGCLFPG
FT CONFLICT 434 478 (IN REF. 3).
FT CONFLICT 478 478 A -> S (IN REF. 2; AAH04229).
FT CONFLICT 417 417 A -> H (IN REF. 3).
FT CONFLICT 433 477 PEGNILDRPFKGFQRRNMTEPERAKFKRKYKVLVEKRAF
FT CONFLICT 434 478 RETQ -> VLTVSCGAPCPVMTSLPVPVPRGYGRHHGCP
FT CONFLICT 478 478 WAGPVGMPRG (IN REF. 5)
FT CONFLICT 478 478 EGNILDRFKSFQRRNMTEPERAKFKRKYKVLVEKRAF
FT CONFLICT 478 478 EIQQL -> RQSHSFETGSAFRRGI (IN REF. 3).
SQ SEQUENCE 478 AA; 54417 MW; 7F18923E348CB52B CRC64;
Query Match 63.5%; Score 33; DB 1; Length 478;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVPXGMSYS 11
DB 239 EVAPAGASYN 248
RESULT 15
BCNS_CLOPE BCNS_CLOPE STANDARD; PRT; 890 AA.
AC P08696;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Bacteriocin BCN5.
GN BCN.
OS Clostridium perfringens.
OG Plasmid pIP404.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RX [1]
RP SEQUENCE FROM N.A.
RC STPAIN=CN50;
RX MEDLINE=88336297; PubMed=2901768;
RA Garnier T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STPAIN=CN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garnier T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium

```

RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene."
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro."
RL Mol. Microbiol. 2:607-614(1988).
CC -!- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -!- INDUCTION: By UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -.
DR EMBL; M32882; AAA98249.1; -.
DR PIR; A30481; A30481.
DR InterPro; IPR000834; Peptidase_M14.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_carboxypept_1.
DR SMART; SM00287; SH3b; 3.
DR Antibiotic; Bacteriocin; Plasmid.
KW DONAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;
SQ
Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Cy 2 EVVPXGMSY 10
Db 170 EVVPGGFTY 178
|||||:|
|

```

Search completed: June 3, 2004, 11:49:53
Job time : 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	39	75.0	1044	Q8DIH0	Q8diH0 synechococc
2	38	73.1	344	Q815A7	Q815a7 bacillus ce
3	38	73.1	367	Q8F8X1	Q8f8x1 rhizobium 1
4	38	73.1	3472	1 Q74056	Q74056 cenarchaeum
5	37	71.2	840	3 Q9URY8	Q9ury8 schizosacch
6	37	71.2	1123	16 Q8EWD4	Q8ewd4 mycoplasma
7	36	69.2	471	11 Q8R126	Q8r126 mus musculus
8	36	69.2	484	11 Q8VD18	Q8vd18 mus musculus
9	36	69.2	484	11 Q8B7X4	Q8b7x4 mus musculus
10	36	69.2	484	11 Q8BK35	Q8bk35 mus musculus
11	36	69.2	559	16 Q839T9	Q839t9 enterococcu
12	36	69.2	1399	16 Q889X7	Q889x7 pseudomonas
13	35	67.3	225	10 Q40129	Q40129 lycopersico
14	35	67.3	245	16 Q7V6Q4	Q7v6q4 prochloroco
15	35	67.3	425	5 Q9XVK4	Q9xvk4 caenorhabdi
16	35	67.3	495	11 Q8C1D7	Q8c1d7 mus musculus

17	35	67.3	555	4 Q7Z6R0	Q7z6r0 homo sapien
18	35	67.3	583	5 Q9BH83	Q9bh83 plasmodium
19	35	67.3	583	5 Q9BHA5	Q9bha5 plasmodium
20	35	67.3	583	5 Q815S7	Q815s7 plasmodium
21	35	67.3	670	11 Q01487	Q01487 rattus ratt
22	35	67.3	747	16 Q8PMI6	Q8pmi6 xanthomonas
23	35	67.3	747	16 Q8PAT2	Q8pat2 xanthomonas
24	35	67.3	749	16 Q9PDM6	Q9pdm6 xylella fas
25	35	67.3	1902	4 Q14122	Q14122 homo sapien
26	34	65.4	156	3 Q13479	Q13479 saccharomyc
27	34	65.4	219	17 Q97182	Q97182 sulfobolus
28	34	65.4	252	17 Q28342	Q28342 archaeoglob
29	34	65.4	290	4 Q96MU1	Q96mu1 homo sapien
30	34	65.4	387	16 Q92MD6	Q92md6 rhizobium m
31	34	65.4	489	4 Q81YM3	Q81ym3 homo sapien
32	34	65.4	541	16 Q98BPS	Q98bps rhizobium 1
33	34	65.4	544	16 Q9PQD2	Q9pqd2 ureaplasma
34	34	65.4	731	16 Q7UWU7	Q7uwu7 rhodopirell
35	34	65.4	842	3 Q9URR4	Q9urra4 penicillium
36	34	65.4	899	16 Q8G4I5	Q8g4i5 bifidobacte
37	34	65.4	1049	16 Q8XT05	Q8xt05 ralstonia s
38	34	65.4	1400	16 Q87KQ5	Q87kq5 vibrio para
39	33	63.5	143	17 Q8TX62	Q8tx62 methanopyru
40	33	63.5	166	16 Q8PPPS	Q8ppps xanthomonas
41	33	63.5	193	2 Q8VUA8	Q8vua8 lactococcus
42	33	63.5	208	2 Q8K7Q4	Q8ktq4 candidatus
43	33	63.5	209	16 Q8RES6	Q8res6 fusobacteri
44	33	63.5	251	13 Q7SY67	Q7sy67 xenopus lae
45	33	63.5	282	16 Q7U552	Q7u552 synechococc

ALIGNMENTS

RESULT 1

Q8DIH0	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
AC	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
DT	01-MAR-2003 (TRENBLrel. 23, Created)		
DT	01-MAR-2003 (TRENBLrel. 23, Last sequence update)		
DT	01-JUN-2003 (TRENBLrel. 24, Last annotation update)		
DE	Multidrug efflux transporter.		
GN	TLL1618.		
OS	Synechococcus elongatus (Thermosynechococcus elongatus).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI_TaxID=32046;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-1;		
RX	MEDLINE=2225144; PubMed=12240834;		
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,		
RA	Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,		
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of the thermophilic cyanobacterium		
RT	Thermosynechococcus elongatus BP-1.";		
RL	DNA Res. 9:123-130(2002).		
DR	EMBL; AP005374; BAC09170.1;		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0005215; F:transporter activity; IEA.		
DR	GO; GO:0006810; P:transport; IEA.		
DR	InterPro; IPR001036; Acriflavin_res.		
DR	InterPro; IPR004784; HAE1.		
DR	Pfam; PF00873; ACR_tran; 1.		
DR	PRINTS; PR00702; ACRIFLAVINRP.		
DR	TIGRFAMs; TIGR00915; 2A0602; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;		

Query Match 75.0%; Score 39; DB 16; Length 1044;

Best local similarity 63.6%; Pred. No. 28;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;


```

QY 1 BEVVPXGMSYS 11
    ||||| :|||
Db 843 BEVLPNGIGYS 853

RESULT 2
Q815A7 Q815A7 PRELIMINARY; PRT; 344 AA.
AC Q815A7;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BC5259.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapatral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Golsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haeelkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyripides N.
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AB017015; AAP12123.1; -.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 344 AA; 39539 MW; C55268ACB7225995 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 344;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
    ||||| :|||
Db 152 BEIAPLGISY 161

RESULT 3
Q98FX1 Q98FX1 PRELIMINARY; PRT; 387 AA.
AC Q98FX1;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hippurate hydrolase.
GN MLR3583.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AF003002; BAB50445.1; -.
DR GO; GO:001678; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase_M20.

Pfam; PF01546; Peptidase_M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
    ||||| :|||
Db 367 DEAIPIHGMSY 376

RESULT 4
O74056 O74056 PRELIMINARY; PRT; 3472 AA.
AC O74056;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
OC Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B;
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL; AF083072; AAC62699.1; -.
DR PIR; T31308; T31308.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006910; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR001680; WD40.
DR SMART; SMO0320; WD40; 2.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Hypothetical protein.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
    ||||| :|||
Db 2294 EDVIPRGISFS 2304

RESULT 5
Q9URY8 Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972h-;
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;

```

IL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

IR EMBL; AL132779; CAB60015.1; -.
IR PIR; T39116; T39116.
IR GeneDB SPombe; SPAC869.05c; -.
IR GO; GO:0016020; C:membrane; IEA.
IR GO; GO:0008271; F:sulfate porter activity; IEA.
IR GO; GO:0008272; P:sulfate transport; IEA.
IR InterPro; IPR002645; STAS.
IR InterPro; IPR001902; Sulpho_transpt.
IR Pfam; PF01740; STAS; 1.
IR TIGRFAMs; TIGR00916; Sulfate transp; 1.
IR PROSITE; PS0801; STAS; 1.
IQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

IY 3 VWPXGMSYS 11

||| ||| |||
135 VWPQMSYA 143

RESULT 6

ID Q8EWD4 PRELIMINARY; PRT; 1123 AA.
AC Q8EWD4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MYPE 2560 paralog, 57%.
EN MYPE2710.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RL MEDLINE=22354719; PubMed=12466555;
LA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
IA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
intracellular bacterial pathogen in humans".
IL Nucleic Acids Res. 30:5293-5300(2002).
IR EMBL; AP004171; BAC44062.1; -.
IR InterPro; IPR008985; ConA like lec.gl.
IR InterPro; IPR007326; Lipoprotein_17.
IR Pfam; PF04200; Lipoprotein_17; 3.
IR Complete proteome.
IQ SEQUENCE 1123 AA; 123636 MW; A4D707330E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
Best Local Similarity 70.0%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

IY 2 EVVPXGMSYS 11

||||| ||| |||
658 EYVPMGLSYS 667

RESULT 7

ID Q8R126 PRELIMINARY; PRT; 471 AA.
AC Q8R126;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
EN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

IX The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
IR EMBL; BC025810; AAH25810.1; -.
IR MGD; MGI:2154441; GLTSCR2.
IR Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;
Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

IY 2 EVVPXGMSYS 11

||||| ||| |||
226 EVIPAGASYN 235

RESULT 8

ID Q8VD18 PRELIMINARY; PRT; 484 AA.
AC Q8VD18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AW556441.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
IR EMBL; BC017637; AAH17637.1; -.
IR MGD; MGI:2154441; GLTSCR2.
SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

IY 2 EVVPXGMSYS 11

||||| ||| |||
239 EVIPAGASYN 248

RESULT 9

ID Q8BTX4 PRELIMINARY; PRT; 484 AA.
AC Q8BTX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
protein.
EN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of

```

RT 60,770 full-length cDNAs.;
RL Nature 420:563-573(2002).
DR EMBL; AK088461; BAC40367.1; -.
DR MGI; MGI:2154441; Gtscr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EBCAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
|||:|:|:|
Db 239 EVIPAGASYN 248

RESULT 10
Q8BK35 PRELIMINARY; PRT; 484 AA.
AC Q8BK35;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE Protein.
DE Gtscr2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK077341; BAC36760.1; -.
DR MGI; MGI:2154441; Gtscr2.
SQ SEQUENCE 484 AA; 55792 MW; EB67949BCB92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
|||:|:|:|
Db 239 EVIPAGASYN 248

RESULT 11
Q839T9 PRELIMINARY; PRT; 559 AA.
AC Q839T9;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Eukaryotic binding protein, putative.
DE EF0663.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22350857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;

```

```

RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016947; AAOV9943.1; -.
DR TIGR; EF0063; -.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR Complete Proteome.
KW SEQUENCE 559 AA; 61476 MW; CC15418D33D53DE7 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 559;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGXMSYS 11
|||:|:|:|
Db 351 LIPEGMSYS 359

RESULT 12
Q889X7 PRELIMINARY; PRT; 1399 AA.
AC Q889X7;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE DNA-directed RNA polymerase, beta subunit.
DE RPOC OR PSPT00620.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016828; AAO54182.1; -.
DR TIGR; PSPT00620; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR DNA-directed RNA polymerase, Complete proteome.
KW SEQUENCE 1399 AA; 154733 MW; 26178D8D53102ADE CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1399;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 10
|||:|:|:|
Db 581 QVVPAGLSY 589

```

```
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGMSYS 11
DB 179 QVVPVGLGYS 188
      :|||:|
      :|||:|

RESULT 15
Q9XVK4
ID Q9XVK4 PRELIMINARY; PRT; 425 AA.
AC Q9XVK4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EXBL; Z81109; CAB03241.1; -.
DR PTR; T24111; T24111.
DR WormPep; R10D12.10; CE12690.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
DB 335 EQVPGGLQY 344
      :|||:|
      :|||:|

Search completed: June 3, 2004, 11:57:32
Job time : 29.8667 secs

Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGMSYS 11
DB 179 QVVPVGLGYS 188
      :|||:|
      :|||:|

RESULT 15
Q9XVK4
ID Q9XVK4 PRELIMINARY; PRT; 425 AA.
AC Q9XVK4;
DT 01-NOV-1999 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein precursor.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=VF36; TISSUE=Pistil;
RX MEDLINE=95375233; PubMed=7647301;
RA Milligan S.B.; Gasser C.S.;
RT "Nature and regulation of pistil-expressed genes in tomato.";
RL Plant Mol. Biol. 28:691-711 (1995).
DR EMBL; U20592; AAA80497.1; -.
DR PIR; S57810; S57810.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PRO0291; KUNITZINHBTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; ST1; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
DB 32 DEVVPGNGTYA 42
      :|||:|
      :|||:|

RESULT 14
Q7V6Q4
ID Q7V6Q4 PRELIMINARY; PRT; 245 AA.
AC Q7V6Q4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospholipid and glycerol acyltransferase (From 'motifs_6.msfl').
RN PMT1092.
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G.; Larimer F.W.; Lamerdin J.; Malfatti S.; Chain P.;
RA Ahlgren N.A.; Arellano A.; Coleman M.; Hauser L.; Hess W.R.;
RA Johnson Z.I.; Land M.; Lindell D.; Post A.P.; Regala W.; Shah M.;
RA Shaw S.L.; Steglich C.; Sullivan M.B.; Ting C.S.; Tolonen A.;
RA Webb E.A.; Zinser E.R.; Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047 (2003).
DR EMBL; BX572098; CAE21267.1; -.
KW Acyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 245 AA; 26907 MW; 106F7C4CB2C6427 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 245;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

W protein - protein search, using sw model

run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-11

Perfect score: 56

Sequence: 1 BEVVPXGMYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	54	96.4	11	5 ABB80523	Abb80523 Hepatitis
2	54	96.4	11	5 ABB80558	Abb80558 Hepatitis
3	54	96.4	11	5 ABB80560	Abb80560 Hepatitis
4	54	96.4	11	5 ABB80527	Abb80527 Hepatitis
5	49	87.5	11	5 ABB80537	Abb80537 Hepatitis
6	49	87.5	11	5 ABB80541	Abb80541 Hepatitis
7	48	85.7	11	5 ABB80546	Abb80546 Hepatitis
8	48	85.7	11	5 ABB80554	Abb80554 Hepatitis
9	48	85.7	11	5 ABB80550	Abb80550 Hepatitis
10	48	85.7	11	5 ABB80555	Abb80555 Hepatitis
11	46	82.1	11	5 ABB80532	Abb80532 Hepatitis
12	46	82.1	11	5 ABB80531	Abb80531 Hepatitis
13	45	80.4	11	5 ABB80525	Abb80525 Hepatitis
14	45	80.4	11	5 ABB80561	Abb80561 Hepatitis
15	45	80.4	11	5 ABB80521	Abb80521 Hepatitis
16	45	80.4	11	5 ABB80522	Abb80522 Hepatitis
17	45	80.4	11	5 ABB80566	Abb80566 Hepatitis
18	45	80.4	11	5 ABB80563	Abb80563 Hepatitis
19	45	80.4	11	5 ABB80565	Abb80565 Hepatitis
20	45	80.4	11	5 ABB80524	Abb80524 Hepatitis
21	45	80.4	11	5 ABB80529	Abb80529 Hepatitis
22	45	80.4	11	5 ABB80567	Abb80567 Hepatitis
23	45	80.4	11	5 ABB80528	Abb80528 Hepatitis
24	45	80.4	11	5 ABB80562	Abb80562 Hepatitis
25	45	80.4	11	5 ABB80559	Abb80559 Hepatitis

26	45	80.4	11	5 ABB80526	Abb80526 Hepatitis
27	45	80.4	11	5 ABB80564	Abb80564 Hepatitis
28	45	80.4	11	5 ABB80568	Abb80568 Hepatitis
29	40	71.4	11	5 ABB80536	Abb80536 Hepatitis
30	40	71.4	11	5 ABB80542	Abb80542 Hepatitis
31	40	71.4	11	5 ABB80543	Abb80543 Hepatitis
32	40	71.4	11	5 ABB80535	Abb80535 Hepatitis
33	40	71.4	11	5 ABB80538	Abb80538 Hepatitis
34	40	71.4	11	5 ABB80540	Abb80540 Hepatitis
35	40	71.4	11	5 ABB80539	Abb80539 Hepatitis
36	39	69.6	11	5 ABB80548	Abb80548 Hepatitis
37	39	69.6	11	5 ABB80549	Abb80549 Hepatitis
38	39	69.6	11	5 ABB80547	Abb80547 Hepatitis
39	39	69.6	11	5 ABB80544	Abb80544 Hepatitis
40	39	69.6	11	5 ABB80556	Abb80556 Hepatitis
41	39	69.6	11	5 ABB80557	Abb80557 Hepatitis
42	39	69.6	11	5 ABB80551	Abb80551 Hepatitis
43	39	69.6	11	5 ABB80553	Abb80553 Hepatitis
44	39	69.6	11	5 ABB80552	Abb80552 Hepatitis
45	39	69.6	11	5 ABB80545	Abb80545 Hepatitis

ALIGNMENTS

RESULT 1

ABB80523

ID ABB80523 standard; peptide; 11 AA.

XX ABB80523;

XX

DT 08-OCT-2002 (first entry)

XX

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.

XX

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with

FT residue 7"

FT

FT Misc-difference 9

FT /note= "D-form residue"

FT

FT Modified-site 11

FT /note= "C-terminal amide"

FT

FN WO200208251-A2.

XX

XX 31-JAN-2002.

XX

XX 19-JUL-2001; 2001WO-US023169.

XX

XX 21-JUL-2000; 2000US-0220101P.

XX

XX (CORV-) CORVAS INT INC.

XX

XX Lim-Wilby M, Levy OE, Brunck TK;

XX

XX WPI; 2002-361643/39.

XX

XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus

PT

PT protease.

XX

XX Claim 17; Page 64; 69pp; English.

XX

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGMHYS 11

RESULT 2
ABB80558
ID ABB80558 standard; peptide; 11 AA.
XX
AC ABB80558;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGMHYS 11

RESULT 3
ABB80560
ID ABB80560 standard; peptide; 11 AA.
XX
AC ABB80560;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGMHYS 11

RESULT 3
ABB80560
ID ABB80560 standard; peptide; 11 AA.
XX
AC ABB80560;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGMHYS 11

2Y 1 BEVVPXGMHYS 11
|||||
2b 1 BEVVPXGMHYS 11

RESULT 4

ABB80527
ID ABB80527 standard; peptide; 11 AA.
XX
AC ABB80527;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 8 residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.

19-JUL-2001; 2001WO-US023169.

21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.

Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C virus
protease.

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 96.4%; Score 54; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
|||||
Db 1 BEVVPXGMHYS 11

RESULT 5

ABB80537
ID ABB80537 standard; peptide; 11 AA.

XX ABB80537;
AC
XX
DT 08-OCT-2002 (first entry)
XX

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.

XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Misc-difference 9 residue 7"

FT Modified-site 11 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX
PN WO200208251-A2.

XX
PD 31-JAN-2002.

19-JUL-2001; 2001WO-US023169.

21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.

Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C virus
protease.

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 87.5%; Score 49; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0069;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
|||||
Db 1 BEVVPXGMHYS 11

RESULT 6

ABB80541
ID ABB80541 standard; peptide; 11 AA.

XX ABB80541;

XX
DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.

XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.
 XX Synthetic.
 OS Key
 XX Location/Qualifiers
 XX Modified-site 1
 XX /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 EN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX protease.
 XX Sequence 11 AA;
 PS Query Match 87.5%; Score 49; DB 5; Length 11;
 XX Best Local Similarity 90.9%; Pred. No. 0.0069;
 XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHYS 11
 DB ||||| |||||
 1 BEVVPXGMHYS 11
 RESULT 7
 ABB80546
 ID ABB80546 standard; peptide; 11 AA.
 AC ABB80546;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX Key Location/Qualifiers
 XX Modified-site 1
 XX /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 EN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX protease.
 XX Sequence 11 AA;
 PS Query Match 87.5%; Score 49; DB 5; Length 11;
 XX Best Local Similarity 90.9%; Pred. No. 0.0069;
 XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHYS 11
 DB ||||| |||||
 1 BEVVPXGMHYS 11
 RESULT 7
 ABB80546
 ID ABB80546 standard; peptide; 11 AA.
 AC ABB80546;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX Key Location/Qualifiers
 XX Modified-site 1
 XX /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 EN 31-JAN-2002.

FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX protease.
 XX Sequence 11 AA;
 PS Query Match 85.7%; Score 48; DB 5; Length 11;
 XX Best Local Similarity 90.9%; Pred. No. 0.011;
 XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHYS 11
 DB ||||| |||||
 1 BEVVPXGMHYS 11
 RESULT 8
 ABB80554
 ID ABB80554 standard; peptide; 11 AA.
 AC ABB80554;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX Key Location/Qualifiers
 XX Modified-site 1
 XX /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 EN 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX protease.
XX Sequence 11 AA;
XX Query Match 85.7%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.011;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 9
ABB0550
ID ABB0550 standard; peptide; 11 AA.
AC ABB0550;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX protease.
XX Sequence 11 AA;
XX Query Match 85.7%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.011;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 10
ABB0555
ID ABB0555 standard; peptide; 11 AA.
XX ABB0555;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

Query Match 85.7%; Score 48; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
Db 1 EEVVPXGMHYS 11

RESULT 11

ABB80532
ID ABB80532 standard; peptide; 11 AA.

XX
AC ABB80532;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
EN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

CC
XX
SQ

pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;
Query Match 82.1%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
Db 1 EEVVPXGMHYS 11

RESULT 12

ABB80531
ID ABB80531 standard; peptide; 11 AA.

XX
AC ABB80531;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX
EN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;
Query Match 82.1%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11
 FT FT /note= "C-terminal amide"
 XX
 FN WO200208251-A2.
 XX
 XX 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Bruck TK;
 PI
 XX WPI; 2002-361643/39.
 DE
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 FT
 XX Claim 17; Page 64; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.042;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHYS 11
 |||||
 Db 1 BEVVPXGMHYS 11

Search completed: June 3, 2004, 11:48:23
 Job time : 45.933 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

WM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	69.6	1037	4	US-09-134-001C-4794
2	37	66.1	856	4	US-08-246-361A-4
3	34	60.7	323	4	US-09-543-681A-7304
4	34	60.7	600	2	US-08-821-119-19
5	34	60.7	600	2	US-08-821-118-2
6	33	58.9	277	4	US-09-252-991A-26615
7	33	58.9	385	4	US-09-252-991A-27834
8	33	58.9	747	4	US-09-724-864-36
9	33	58.9	3472	4	US-08-460-020-4
10	32	57.1	70	4	US-09-134-001C-3950
11	32	57.1	101	4	US-09-621-976-6096
12	32	57.1	102	2	US-08-580-988A-23
13	32	57.1	126	2	US-08-879-995A-3
14	32	57.1	126	3	US-09-215-096-3
15	32	57.1	152	2	US-08-460-694-4
16	32	57.1	152	3	US-08-460-744-4
17	32	57.1	152	3	US-07-667-711B-4
18	32	57.1	173	1	US-08-193-977-7
19	32	57.1	189	2	US-08-464-517-21
20	32	57.1	189	2	US-08-246-361A-21
21	32	57.1	189	3	US-08-463-772-21
22	32	57.1	189	5	PCT-US93-05000-21
23	32	57.1	236	2	US-08-464-517-22
24	32	57.1	236	2	US-08-246-361A-22
25	32	57.1	236	3	US-08-463-772-22
26	32	57.1	236	5	PCT-US93-05000-22
27	32	57.1	280	2	US-08-464-517-6

Query Match	Score	DB 4	Length	1037	Indels	Gaps
Best Local Similarity	69.6%	63.6%	2	Mismatches	2	Indels
Matches	7	Conservative	1	BEVVPXGMHYS 11		
Qy	1	BEVVPXGMHYS 11				
Db	199	KEVVSNGHYS 209				

Query Match	Score	DB 4	Length	1037	Indels	Gaps
Best Local Similarity	69.6%	63.6%	2	Mismatches	2	Indels
Matches	7	Conservative	1	BEVVPXGMHYS 11		
Qy	1	BEVVPXGMHYS 11				
Db	199	KEVVSNGHYS 209				

Query Match	Score	DB 4	Length	1037	Indels	Gaps
Best Local Similarity	69.6%	63.6%	2	Mismatches	2	Indels
Matches	7	Conservative	1	BEVVPXGMHYS 11		
Qy	1	BEVVPXGMHYS 11				
Db	199	KEVVSNGHYS 209				

Sequence 6, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 23, Appli
Sequence 6, Appli
Sequence 23, Appli
Sequence 23, Appli
Sequence 23, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 19, Appli
Sequence 20, Appli
Sequence 19, Appli
Sequence 20, Appli
Sequence 19, Appli
Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-4794
; Sequence 4794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4794
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

RESULT 2
US-09-252-991A-21444
; Sequence 21444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21444
; LENGTH: 856
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 66.1%; Score 37; DB 4; Length 856;
Best Local Similarity 70.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVPGXGMHY 10
Db 64 EAVVPGGEHY 73

RESULT 3

US-09-543-681A-7304
; Sequence 7304; Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543.681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 7304

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-7304

Query Match 60.7%; Score 34; DB 4; Length 323;

Best Local Similarity 55.6%; Pred. No. 63;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPGXGMHY 10

Db 75 DVCPAGVHY 83

RESULT 4

US-08-821-119-19

; Sequence 19; Application US/08821119

; Patent No. 5821104

; GENERAL INFORMATION:

; APPLICANT: Holm, Kaj Andre

; APPLICANT: Rasmussen, Grethe

; APPLICANT: Halkier, Torben

; APPLICANT: Lehmebeck, Jan

; TITLE OF INVENTION: Tripeptidyl Aminopeptidase

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 58211040 No. 5821104disk of No. 5821104th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: Fast-SEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/821.119

; FILING DATE: 19-MAR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4107.204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 600 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-08-821-119-19

Query Match 60.7%; Score 34; DB 2; Length 600;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VPGXGMHYS 11

Db 31 VPGXGMHYS 38

RESULT 5

US-08-821-118-2

; Sequence 2; Application US/08821118

; Patent No. 5989889

; GENERAL INFORMATION:

; APPLICANT: Rey, Michael

; APPLICANT: Golightly, Elizabeth

; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES

; TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE

; TITLE OF INVENTION: ACTIVITY

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 59898890 No. 5989889disk of No. 5989889th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/821.118

; FILING DATE: 19-MAR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4107.400-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 600 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-08-821-118-2

```

Db      31 VPKGMHYS 38

RESULT 6
US-09-252-991A-26615
; Sequence 26615, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26615
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match      58.9%; Score 33; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 BEVVPKGMHYS 11
      |||||
Db      48 BEVVPKGMHYS 58

RESULT 7
US-09-252-991A-27834
; Sequence 27834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27834
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27834

Query Match      58.9%; Score 33; DB 4; Length 385;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 BEVVPKGMHYS 10
      |||||
Db      201 EILPAALHY 209

RESULT 8
US-09-724-864-36
; Sequence 36, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed

; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 11000.1050U
; CURRENT APPLICATION NUMBER: US/09/724,864
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Rat
US-09-724-864-36

Query Match      58.9%; Score 33; DB 4; Length 747;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      5 PXXGMHYS 11
      |||||
Db      627 FGGLHYS 633

RESULT 9
US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match      58.9%; Score 33; DB 4; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.4e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVVPKGMHYS 11
      |||||
Db      2294 EDVIPRGISFS 2304

RESULT 10
US-09-134-001C-3950
; Sequence 3950, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3950
; LENGTH: 70
; TYPE: PRT

```

; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3950

Query Match 57.1%; Score 32; DB 4; Length 70;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VPXGMHYS 11
Db 36 MPKGFHYS 43

RESULT 11

US-09-621-976-6096
; Sequence 6096, Application US/09621976
; Patent No. 8639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.PM
; SEQ ID NO 6096
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 96
; OTHER INFORMATION: Xaa = * ,Ala,Glu,Gly,Ile,Lys,Leu,Arg,Ser,Thr,Val

US-09-621-976-6096

Query Match 57.1%; Score 32; DB 4; Length 101;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PXGMHY 10
Db 40 PRGMHY 45

RESULT 12

US-08-580-988A-23
; Sequence 23, Application US/08580988A
; Patent No. 5856161

; GENERAL INFORMATION:
; APPLICANT: Aggarwal et al. Tumor Necrosis Factor
; TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
; TITLE OF INVENTION: For its Use
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,988A
; FILING DATE: January 3, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5721CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: no
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-580-988A-23

Query Match 57.1%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
Db 24 BEVFPPLAMNY 33

RESULT 13

US-08-879-995A-3
; Sequence 3, Application US/08879995A
; Patent No. 5985606

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Kaber, Matthew R.
; TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,995A
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0326 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
US-08-879-995A-3

Query Match 57.1%; Score 32; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 9
DB 28 EQVPGGHH 36

RESULT 14

US-09-215-096-3

Sequence 3, Application US/09215096

Patent No. 6008194

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Kaser, Matthew R.

TITLE OF INVENTION: HUMAN PREPROTACHYKININ B

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/215,096

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/879,995

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0326 US

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 163590

US-09-215-096-3

Query Match 57.1%; Score 32; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 9
DB 28 EQVPGGHH 36

RESULT 15

US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5858655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRADI Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,694

FILING DATE: 02-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: McConathy, Evelyn H.

REGISTRATION NUMBER: 35,279

REFERENCE/DOCKET NUMBER: 0609.4070002

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 152 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-460-694-4

Query Match 57.1%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 10
DB 20 EEVFPPLANNY 29

Search completed: June 3, 2004, 12:03:07
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-11

Perfect score: 56

Sequence: 1 EEVVPXGMHYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	96.4	11	12	US-09-909-164-7
2	54	96.4	11	12	US-09-909-164-11
3	49	87.5	11	12	US-09-909-164-21
4	49	87.5	11	12	US-09-909-164-25
5	48	85.7	11	12	US-09-909-164-30
6	48	85.7	11	12	US-09-909-164-34
7	48	85.7	11	12	US-09-909-164-38
8	48	85.7	11	12	US-09-909-164-39
9	48	85.7	11	12	US-09-909-164-42
10	48	85.7	11	12	US-09-909-164-44
11	46	82.1	11	12	US-09-909-164-15
12	46	82.1	11	12	US-09-909-164-16
13	45	80.4	11	12	US-09-909-164-5
14	45	80.4	11	12	US-09-909-164-6
15	45	80.4	11	12	US-09-909-164-8

Sequence 9, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 47, Appl
Sequence 48, Appl
Sequence 49, Appl
Sequence 50, Appl
Sequence 51, Appl
Sequence 52, Appl
Sequence 19, Appl
Sequence 20, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 24, Appl
Sequence 26, Appl
Sequence 27, Appl
Sequence 28, Appl
Sequence 29, Appl
Sequence 31, Appl
Sequence 32, Appl
Sequence 33, Appl
Sequence 35, Appl
Sequence 36, Appl
Sequence 37, Appl
Sequence 40, Appl
Sequence 41, Appl
Sequence 43, Appl
Sequence 45, Appl
Sequence 46, Appl

ALIGNMENTS

RESULT 1

US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)

```

; OTHER INFORMATION: AMIDATION
US-09-909-164-7

Query Match      96.4%; Score 54; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 2
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-11

Query Match      96.4%; Score 54; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 3
US-09-909-164-21
; Sequence 21, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US

```

```

; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-21

Query Match      87.5%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0067;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMHYS 11
      |||||
Db      1 EEVVPXGMHYS 11

RESULT 4
US-09-909-164-25
; Sequence 25, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

```

OTHER INFORMATION: norvaline-(CO)
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
US-09-909-164-25

Query Match 87.5%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0067;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGQHS 11

RESULT 5

US-09-909-164-30
Sequence 30, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 30
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
NAME/KEY: MISC FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
US-09-909-164-34

Query Match 85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGSHYS 11

RESULT 6

US-09-909-164-34
Sequence 34, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
US-09-909-164-30

Query Match 85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGTHYS 11

RESULT 7

US-09-909-164-38
Sequence 38, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 34
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
US-09-909-164-34

Query Match 85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGSHYS 11

RESULT 7

US-09-909-164-38
Sequence 38, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)

RESULT 7

US-09-909-164-38
Sequence 38, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)

```

; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-38
Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY 1 EEVVPXGMHYS 11
   |||||
Db 1 EEVVPXGSHYS 11

RESULT 8
US-09-909-164-39
; Sequence 39, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 39
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met(O)
US-09-909-164-42
Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
   |||||
Db 1 EEVVPXGSHYS 11

RESULT 10
US-09-909-164-44
; Sequence 44, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 44
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-39
Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY 1 EEVVPXGMHYS 11
   |||||
Db 1 EEVVPXGSHYS 11

RESULT 9
US-09-909-164-42
; Sequence 42, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita

```

```
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; US-09-909-164-44
```

```
Query Match      85.7%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.01;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGHHYS 11
   |||||
DB 1 EEVVPXGHHYS 11
```

```
RESULT 11
US-09-909-164-15
; Sequence 15, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence X
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; US-09-909-164-15
```

```
Query Match      82.1%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGHHYS 11
   |||||
DB 1 EEVVPXGHHYS 11
```

```
RESULT 12
US-09-909-164-16
; Sequence 16, Application US/09909164
```

```
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence X
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-16
```

```
Query Match      82.1%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGHHYS 11
   |||||
DB 1 EEVVPXGHHYS 11
```

```
RESULT 13
US-09-909-164-5
; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence X
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
```

OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-5

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 14

US-09-909-164-6
Sequence 6, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 6
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-6

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 15

US-09-909-164-8
Sequence 8, Application US/09909164

Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 8
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-8

Query Match 80.4%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
| | | | | | | | | |
Db 1 EEVVPXGMDYS 11

Search completed: June 3, 2004, 12:57:15
Job time : 33.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	66.1	308	2 A72207	ftsH proteinase ac
2	37	66.1	1057	2 F89892	carbamoyl-phosphat
3	36	64.3	102	2 A42452	V1 protein - tobac
4	36	64.3	252	2 A82001	hypothetical prote
5	36	64.3	460	2 S82046	hypothetical prote
6	36	64.3	743	2 S81433	hypothetical prote
7	35	62.5	156	2 D82618	conserved hypothet
8	35	62.5	233	2 T02590	DNA binding protei
9	35	62.5	311	2 H69194	GMP synthetase, su
10	35	62.5	425	2 T24111	hypothetical prote
11	35	62.5	510	2 G86430	T518.1 protein - A
12	34	60.7	264	2 G92117	diphthine synthase
13	34	60.7	279	2 C75538	hypothetical prote
14	34	60.7	350	2 B75478	3-dehydroquinatase
15	34	60.7	355	2 T35025	probable DNA ligas
16	34	60.7	360	2 E69086	cell division prot
17	34	60.7	425	2 C83903	hypothetical prote
18	34	60.7	426	2 S58132	Slg1 protein precu
19	34	60.7	495	2 T28717	hypothetical prote
20	34	60.7	1028	2 A32866	ATP-dependent DNA
21	33	58.9	156	2 S54619	hypothetical prote
22	33	58.9	367	2 E83607	polyamine transpor
23	33	58.9	441	2 G82253	conserved hypothet
24	33	58.9	466	2 H71542	probable amino aci
25	33	58.9	466	2 H81697	amino acid antipor
26	33	58.9	487	2 S55811	finger protein (cl
27	33	58.9	514	1 HQVULS	cytochrome-c3 hydr
28	33	58.9	534	2 A69284	coenzyme F420-hydr
29	33	58.9	545	2 T08564	hypothetical prote

30 33 58.9 627 2 A69663 DNA mismatch repai
31 33 58.9 716 1 JC5061 macrophage-stimula
32 33 58.9 1257 2 S44754 Cl4B9.8 protein -
33 33 58.9 1396 2 S36851 L-shaped tail fi
34 33 58.9 3472 2 T31308 hypothetical 167K
35 32 57.1 126 2 A25905 tachykinin B precu
36 32 57.1 197 2 D71840 heme exporter prot
37 32 57.1 225 2 S57810 hypothetical prote
38 32 57.1 233 2 E97120 ribosomal protein
39 32 57.1 267 2 T07215 ribosomal protein
40 32 57.1 270 2 C95881 probable ThuaA prot
41 32 57.1 283 2 T25737 hypothetical prote
42 32 57.1 288 2 JC4011 cycclin D2 - rat
43 32 57.1 288 2 I58372 cycclin D2 - rat
44 32 57.1 289 2 A41984 cycclin D2 - mouse
45 32 57.1 289 2 A42822 cycclin D2 - human

ALIGNMENTS

RESULT 1

A72207
ftsH proteinase activity modulator HflK - Thermotoga maritima (strain MS88)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72207
R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.N.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D., C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; PMID:99287316; PMID:10360571
A:Accession: A72207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <ARN>
A:Cross-references: GB:AE001819; GB:AE000512; NID:g4982396; PIDN:AAD36885.1; PID:g498240
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1822
C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 56.1%; Score 37; DB 2; Length 308;
Best Local Similarity 75.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
||| |
Db 41 VVPSGIHY 48

RESULT 2

F89892
carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89892
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; PMID:21311952; PMID:11418146
A:Accession: F89892
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1057 <KUR>
A:Cross-references: GB:BA000018; PID:g13701002; PIDN:BA842298.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: PyrB8
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain, biotin


```

Query Match      66.1%; Score 37; DB 2; Length 1057;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMHYS 11
    ||| |||
Db 190 EIVSNGLHYS 199

RESULT 3
A:Reference number: A42452
A:Accession: A42452
A:Molecule type: DNA
C:Species: tobacco yellow dwarf virus (strain Australia)
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <WOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match      64.3%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMHYS 11
    ||| |||
Db 7 QVVPISGINS 16

RESULT 4
A:Reference number: AE2001
A:Accession: AE2001
A:Molecule type: DNA
C:Species: Nostoc sp. PCC 7120
C:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE2001
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA577929.1; PID:g17135383; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1563

Query Match      64.3%; Score 36; DB 2; Length 252;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
    ||| |||
Db 235 EMIVPAGLHF 244

RESULT 5
A:Reference number: S69046
A:Accession: S69046
A:Molecule type: DNA
C:Species: Saccharomyces cerevisiae
C>Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 17-Mar-2000
C:Accession: S69046
R:Hall, J.; dePaolo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

Query Match      64.3%; Score 36; DB 2; Length 460;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
    ||| |||
Db 85 IVPLGLHY 92

RESULT 6
A:Reference number: S38143
A:Accession: S38143
A:Molecule type: DNA
C:Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-Apr-2002
C:Accession: S38143
R:van Vliet-Reedijk, J.C.; Pianta, R.J.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38130
A:Accession: S38143
A:Molecule type: DNA
A:Residues: 1-743 <VAN>
A:Cross-references: EMBL:228292; NID:g486536; PIDN:CAA82146.1; PID:g486537; MIPS:YKR067
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GPT2
A:Cross-references: SGD:S0001775
A:Map position: 11R
C:Keywords: transmembrane protein

Query Match      64.3%; Score 36; DB 2; Length 743;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
    ||| |||
Db 294 VVPCGLHY 301

RESULT 7
A:Reference number: D82618
A:Accession: D82618
A:Molecule type: DNA
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82618
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82618
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <SIM>
A:Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GN00
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canagar, L.E.A.; Carraro, D.M.; Carrer,
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieser, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,

```

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF1950

Query Match 62.5%; Score 35; DB 2; Length 156;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 0; Gaps 0;

QY 1 BEVVPXGNH 9
DB 119 BEILPQGVH 127
||| |||

RESULT 8
T02590
DNA binding protein EREBP-2 - common tobacco
A;Species: Nicotiana tabacum (common tobacco)
A;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
A;Accession: T02590
A;Ohme-Takegi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi
A;Reference number: 214671; MUID: 95276459; PMID: 7756828
A;Accession: T02590
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-233 <OHN>
A;Cross-references: EMBL:D38126; NID:9790362; PIDN:BA07324.1; PID:g1208498
A;Experimental source: strain BY4; tissue-type leaf

Query Match 62.5%; Score 35; DB 2; Length 233;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 1 BEVVPXGNH 10
DB 90 QAVVPKGRHY 99
||| |||

RESULT 9
H59194
GMP synthetase, subunit B - Methanobacterium thermoautotrophicum (strain Delta H)
A;Species: Methanobacterium thermoautotrophicum
A;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
A;Accession: H59194
A;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Oiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A;Reference number: A69000; MUID: 98037514; PMID: 9371463
A;Accession: H59194
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-311 <MTH>
A;Cross-references: GB:A5000850; GB:A5000666; NID:g2621794; PIDN:AA85215.1; PID:g262179
A;Experimental source: strain Delta H
A;Genetics:
A;Gene: MTH710
A;Start codon: GTG

Query Match 62.5%; Score 35; DB 2; Length 311;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGNH 11
||| |||

DB 219 EEVVEGLHES 229
RESULT 10
T24111
Hypothetical protein R10D12.10 - Caenorhabditis elegans
A;Species: Caenorhabditis elegans
A;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A;Accession: T24111
A;Percy, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19842
A;Accession: T24111
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A;Experimental source: clone R10D12
A;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 62.5%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
DB 335 EQIVPGGLQY 344
||| |||

RESULT 11
G86430
1518.1 protein - Arabidopsis thaliana
A;Species: Arabidopsis thaliana (mouse-ear cress)
A;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
A;Accession: G86430
A;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID: 21016719; PMID: 11130712
A;Accession: G86430
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-510 <SPO>
A;Cross-references: GB:AE005172; NID:g4587512; PIDN:AA25743.1; GSPDB:GN00141
A;Genetics:
A;Map position: 1
A;Superfamily: hexose phosphate transport protein ubPT

Query Match 62.5%; Score 35; DB 2; Length 510;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
DB 12 BEVKKPGIHF 21
||| |||

RESULT 12
G69117
diphthine synthase - Methanobacterium thermoautotrophicum (strain Delta H)
A;Species: Methanobacterium thermoautotrophicum
A;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 25-Aug-2003
A;Accession: G69117

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funcd
 A;Reference number: A69000; MUID:198037314; PMID:9371463
 A;Accession: G69117
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-264 <MTH>
 A;Cross-references: GB:AE0000940; GB:AE0000666; NID:G2623011; PIDN:AAB86340.1; PID:G262301
 A;Experimental source: strain Delta H
 C;Genetics:
 A;Gene: MTH1874
 C;Superfamily: diphthamide biosynthesis methyltransferase

Query Match 60.7%; Score 34; DB 2; Length 264;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMHY 10
 |||||
 Db 235 VVPAGLHF 242

RESULT 13
 C75538
 Hypothetical protein - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C;Accession: C75538
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: C75538
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-279 <WHI>
 A;Cross-references: GB:AE001889; GB:AE000513; NID:G6457944; PIDN:AAF09867.1; PID:G645795
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0271
 A;Map position: 1
 C;Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 60.7%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 VVPXGMHY 11
 |||||
 Db 100 VPLGRHYS 107

RESULT 14
 B75478
 3-dehydroquinase synthase - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C;Accession: B75478
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: B75478
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-350 <WHI>

A;Cross-references: GB:AE001932; GB:AE000513; NID:G6458481; PIDN:AAF10353.1; PID:G64584
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0777
 A;Map position: 1
 C;Superfamily: 3-dehydroquinase synthase; 3-dehydroquinase synthase homology

Query Match 60.7%; Score 34; DB 2; Length 350;
 Best Local Similarity 60.0%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EYVPGMHYS 11
 |||||
 Db 252 EAVAYGMHYA 261

RESULT 15
 T35025
 Probable DNA ligase - Streptomyces coelicolor
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C;Accession: T35025
 R;Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A;Reference number: Z21565
 A;Accession: T35025
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-355 <SEE>
 A;Cross-references: EMBL:AL079355; PIDN:CAB45581.1; GSPDB:GN000070; SCOEDB:SC4C6.17C
 A;Experimental source: strain A3(2)
 C;Genetics:
 A;Gene: SCOEDB:SC4C6.17C

Query Match 60.7%; Score 34; DB 2; Length 355;
 Best Local Similarity 71.4%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 VVPXGMHY 10
 |||||
 Db 20 IPPGGMHY 26

Search completed: June 3, 2004, 12:00:00
 Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	69.6	1057	1	Q8CPJ4 staphylococ
2	38	67.9	1058	1	Q8R86 fuscobacteri
3	37	66.1	426	1	Q87X9 vibrio para
4	37	66.1	1057	1	Q99R5 staphylococ
5	37	66.1	1057	1	P58940 staphylococ
6	36	64.3	102	1	P31619 tobacco yel
7	36	64.3	460	1	Q03010 saccharomyc
8	36	64.3	743	1	P36148 saccharomyc
9	35	62.5	227	1	O3586 mesocricetu
10	35	62.5	308	1	O26806 methanobact
11	34	60.7	264	1	Q7902 methanobact
12	34	60.7	426	1	Q99158 yarrowia li
13	33	58.9	441	1	P46231 vibrio para
14	33	58.9	513	1	P13065 desulfovibr
15	33	58.9	627	1	P49850 bacillus su
16	33	58.9	1188	1	P34335 caenorhabdi
17	33	58.9	1396	1	P13390 bacterioph
18	32	57.1	126	1	P08858 bos taurus
19	32	57.1	153	1	O76217 anopheles g
20	32	57.1	212	1	Q878w6 vibrio para
21	32	57.1	233	1	Q97166 clostridium
22	32	57.1	267	1	P56351 chlorella v
23	32	57.1	288	1	Q04827 rattus norv
24	32	57.1	289	1	P30280 mus musculus
25	32	57.1	289	1	Q90459 brachydanio
26	32	57.1	291	1	P50755 xenopus lae
27	32	57.1	291	1	P49706 gallus gall
28	32	57.1	291	1	P53782 xenopus lae
29	32	57.1	292	1	P55169 gallus gall
30	32	57.1	292	1	CD1_CHICK
31	32	57.1	292	1	CD3_HUMAN
32	32	57.1	295	1	P24385 homo sapien
33	32	57.1	295	1	CD1_MOUSE

34	32	57.1	295	1	CGD1_RAT
35	32	57.1	341	1	HYPE_AZOVI
36	32	57.1	353	1	T2BA_BACAR
37	32	57.1	573	1	SUOX_DROME
38	32	57.1	578	1	MDLB_BUCBP
39	32	57.1	759	1	SCT1_YEAST
40	32	57.1	877	1	SULT1_SCHPO
41	32	57.1	1401	1	RPOC_VIBCH
42	32	57.1	2717	1	ZEP1_HUMAN
43	31.5	56.2	847	1	CD22_HUMAN
44	31	55.4	124	1	REV_SIVCZ
45	31	55.4	130	1	SZ05_RAT

ALIGNMENTS

RESULT 1
CARB_STAEP STANDARD; PRT; 1057 AA.
AC Q8CPJ4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN CARB OR SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J., Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z., Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RA "Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228).";
RT Mol. Microbiol. 49:1577-1593 (2003).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -1- SIMILARITY: Belongs to the carb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; A016746; AAO04476.1; -.
CC HAMAP; MF 01210; -; 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR004362; MGS-like.
DR InterPro; IPR000169; SHProt_acsite.
DR Pfam; PF00289; CPhase_L_chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPhase.

```
DR TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7DBD8B1CAB59 CRC64;

Query Match 69.6%; Score 39; DB 1; Length 1057;
Best Local Similarity 63.8%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMHYS 11
DB 189 KEVVSNGLHYS 199
:|||||:|||||

RESULT 2
CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
DE CARB OR FNO422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium
RA nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -|- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -|- PATHWAY: Arginine biosynthesis;
CC -|- PATHWAY: Pyrimidine biosynthesis; first step.
CC -|- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -|- SIMILARITY: Belongs to the carb family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC EMBL; AE010554; AAL94625.1; ALT_INIT.
DR HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 2.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 67.9%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMHYS 11
DB 190 EIVPNGLHYS 199
:|||||:|||||

RESULT 3
AROA_VIBPA STANDARD; PRT; 426 AA.
AC Q87QX9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR VP1020.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RA distinct from that of V. cholerae."
RL Lancet 361:743-749(2003).
CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -|- SUBUNIT: Monomer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
```

```

CC -!- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005076; BACS9283.1; -
CC HAMAP; MF 0210; -; 1.
CC InterPro; IPR001986; EPSP synth.
CC Pfam; PF00275; EPSP synthase; 1.
CC PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
CC Aromatic amino acid biosynthesis; Transferase; Complete proteome.
CC KW SEQUENCE 426 AA; 46094 MW; 373D39C5BA1F70F CRC64;
CC
CC Query Match 66.1%; Score 37; DB 1; Length 426;
CC Best Local Similarity 60.0%; Pred. No. 7.3;
CC Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC 2Y 1 EVVFXGNHY 10
CC 223 EFVIPAGQHY 232
CC
CC Db
CC
CC RESULT 4
CC CARB STAAW STANDARD; PRT; 1057 AA.
CC ID CARB STAAW STANDARD; PRT; 1057 AA.
CC AC Q99UR5;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
CC phosphate synthetase ammonia chain).
CC GN CARB OR PYRAB OR SAV1203 OR SA1046.
CC OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
CC Staphylococcus aureus (strain N315).
CC OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
CC OX NCBI_TaxID=158878, 158879;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=Mu50 / ATCC 700699, and N315;
CC MEDLINE=21311952; PubMed=11418146;
CC Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
CC Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
CC Xananori M., Matsumaru H., Maruyama A., Murakami H., Hasevama A.,
CC Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue K.-I., Kaito C.,
CC Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
CC Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
CC Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
CC "Whole genome sequencing of methicillin-resistant Staphylococcus
CC aureus";
CC Lancet 357:1225-1240 (2001).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003361; BAB57365.1; -
CC EMBL; AP003132; BAB42298.1; -
CC PIR; F89892; F89892.
CC HSSP; P00968; 1CS0.
CC HAMAP; MF 01210; -; 1.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPase_L_D2.
CC InterPro; IPR005479; CPase_L_D3.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF00289; CPase_L_Chain; 2.
CC Pfam; PF02786; CPase_L_D2; 2.
CC Pfam; PF02787; CPase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPASE.
CC TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
CC PROSITE; PS00866; CPASE_1; 2.
CC PROSITE; PS00867; CPASE_2; 2.
CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
CC ATP-binding; Manganese; Complete proteome.
CC KW CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
CC FT DOMAIN 1 401 OLIGOMERIZATION DOMAIN.
CC FT DOMAIN 402 546 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
CC FT DOMAIN 547 929 ALLOSTERIC DOMAIN.
CC FT DOMAIN 930 1057
CC FT REPEAT 1 546
CC FT REPEAT 547 1057
CC FT NP_BIND 153 210 ATP (POTENTIAL).
CC FT NP_BIND 302 352 ATP (POTENTIAL).
CC FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
CC FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
CC FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
CC FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
CC FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
CC SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;
CC
CC Query Match 66.1%; Score 37; DB 1; Length 1057;
CC Best Local Similarity 60.0%; Pred. No. 19;
CC Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 2 EVVFXGNHY 11
CC 190 EIVSNGLHY 199
CC
CC Db
CC
CC RESULT 5
CC CARB STAAW STANDARD; PRT; 1057 AA.
CC ID CARB STAAW STANDARD; PRT; 1057 AA.
CC AC P59940;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
CC phosphate synthetase ammonia chain).
CC DE phosphate synthetase ammonia chain).
CC GN CARB OR PYRAB OR MW1086.
CC OS Staphylococcus aureus (strain MW2).
CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
CC OX NCBI_TaxID=196620;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=22040717; PubMed=12044378;
CC Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
CC Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
CC Yamamoto K., Hiramatsu K.;
CC "Genome and virulence determinants of high virulence community-
CC acquired MRSA";
CC Lancet 359:1819-1827 (2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.

```

CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (by
CC similarity).
CC
CC -!- SIMILARITY: Belongs to the carB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP004825; BAB94951.1; -.
CC HMAP; MF 01210; -; 1
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF00289; CPase_L_Chain; 2.
CC Pfam; PF02786; CPase_L_D2; 2.
CC Pfam; PF02787; CPase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPASASE.
CC TIGRFAMs; TIGR01369; CPaseII_lig; 1.
CC PROSITE; PS00866; CPASE_1; 2.
CC PROSITE; PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9C6F152 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPGMHYS 11
Db 190 EIVSNGLHYS 199
|:|:|:|:|:|:|
|:|:|:|:|:|:|
RESULT 6
Y1LK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;

RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RL Virology 187:633-642(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81103; AAA47947.1; -.
CC PIR; A42452; A42452.
CC InterPro; IPR002621; Gemini_mov.
CC Pfam; PF01708; Gemini_mov; 1.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF5B67 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPGMHYS 11
Db 7 QVVPFGINYS 16
|:|:|:|:|:|:|
|:|:|:|:|:|:|
RESULT 7
UMEL YEAST STANDARD; PRT; 460 AA.
AC Q03010; P87330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Meiosis negative regulator UME1.
GN UME1 OR WTM3 OR YPL139C OR LPI7C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364A;
RA Mallory M.J., Strich R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung B., Churcher C.M., Coester F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Patel F.X., Pohl T.M., Furnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Uristarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W., Zollner A., Vo D.H., Hani J.;
RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RT Nature 387:103-105(1997).
CC -!- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -!- SIMILARITY: Contains 4 WD repeats.
CC -!- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U10280; AAB40937.1; -;
CC EMBL; U43703; AAB68221.1; -;
CC PIR; S69046; S69046.
CC Germline; 144121; -;
CC TRANSFAC; T04309; -;
CC SGD; S0006060; UNE1.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0003714; F:transcription co-repressor activity; IDA.
CC GO; GO:0040020; P:regulation of meiosis; IGI.
CC InterPro; IPR001860; WD40.
CC Pfam; PF00400; WD40; 3.
CC SMART; SM00320; WD40; 4.
CC PROSITE; PS00678; WD REPEATS 1; FALSE NEG.
CC PROSITE; PS00082; WD REPEATS 2; FALSE NEG.
CC PROSITE; PS50294; WD REPEATS REGION; FALSE NEG.
CC Transcription regulation; Meiosis; Repeat; WD repeat.
CC REPEAT 233 271 WD 1.
CC REPEAT 276 316 WD 2.
CC REPEAT 339 379 WD 3.
CC REPEAT 411 451 WD 4.
CC SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;
Query Match 64.3%; Score 36; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
:|:|:|:
DB 85 IVPLGLHY 92

RESULT 8
ID YK47 YEAST STANDARD; PRT; 743 AA.
AC P36148.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCP1-MET1 intergenic region.
GN YK067W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA van Vliet-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: STRONG, TO YEAST YB011W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z28292; CAA82146.1; -;
CC PIR; S38143; S38143.
CC Germline; 140046; -;
CC SGD; S0001775; GPT2.

DR GO; GO:0005783; C:endoplasmic reticulum; IDA.
DR GO; GO:0004366; F:glycerol-3-phosphate O-acyltransferase acti...; IDA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IDA.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; Pfam; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 31 55 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.
FT TRANSMEM 502 524 POTENTIAL.
FT TRANSMEM 539 555 POTENTIAL.
SQ SEQUENCE 743 AA; 83644 MW; 84B9948E56B82F15 CRC64;
Query Match 64.3%; Score 36; DB 1; Length 743;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
:|:|:|:
DB 294 VVPCGLHY 301

RESULT 9
ID I1_MESAU STANDARD; PRT; 227 AA.
AC O35586;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Isopentenyl-diphosphate delta-isomerase 1 (EC 5.3.3.2) (IPP isomerase
DE 1) (Isopentenyl pyrophosphate isomerase 1) (IPPI1).
GN IDI1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373600; PubMed=9228075;
RA Paton V.G., Shackelford J.E., Krisans S.K.;
RT "Cloning and subcellular localization of hamster and rat isopentenyl
RT diphosphate dimethylallyl diphosphate isomerase. A PTS1 motif targets
RT the enzyme to peroxisomes."
RL J. Biol. Chem. 272:18945-18950(1997).
CC -!- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
CC HOMALLYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS HIGHLY
CC ELECTROPHILIC ALLYLIC ISOMER, DIMETHYLLALLYL DIPHOSPHATE (DMAPP).
CC -!- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
CC diphosphate.
CC -!- COFACTOR: Magnesium.
CC -!- PATHWAY: Isoprenoid biosynthetic pathway whose end products
CC include dolichols, vitamins A, D, E, and K, steroid hormones,
CC carotenoids bile acids and cholesterol.
CC -!- SUBCELLULAR LOCATION: Peroxisomal.
CC -!- SIMILARITY: Belongs to the IPP isomerase type 1 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF003836; AAC53283.1; -;
CC InterPro; IPR002667; IPP isomerase.
CC Pfam; PF00293; NUDIX; 1.
CC Germline; PD004109; IPP isomerase; 1.
CC Carotenoid biosynthesis; Cholesterol biosynthesis;
CC Isoprene biosynthesis; Sterol biosynthesis; Isomerase; Peroxisome;
CC Magnesium.

Query Match 62.5%; Score 35; DB 1; Length 300;
Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE SLS1 protein precursor.
 SN SLS1.
 CS Yarrowia lipolytica (Candida lipolytica).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Dipodascaceae; Yarrowia.
 CX NCBI_TaxID=4952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20460 / W29;
 RX MEDLINE=96216076; PubMed=8662639;
 RA Boiarane A., Beckerich J.-M., Gaillardin C.; is involved in the
 RT protein translocation process in the yeast Yarrowia lipolytica.";
 RL J. Biol. Chem. 271:11668-11675 (1996).
 CC -!- FUNCTION: Involved in the protein translocation process. May
 CC interact directly with translocating polypeptides to facilitate
 CC their transfer and/or help their folding in the ER. It is not
 CC required for viability but is essential for optimal growth at
 CC elevated temperatures.
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z50154; CAA90516.1; -.
 DR PIR; S58132; S58132.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000886; ER target S.
 DR PROSITE; PS00014; ER_TARGET; 1.
 KW Endoplasmic reticulum; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 426 SLS1 PROTEIN.
 FT SITE 423 426 PREVENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 426 AA; 47201 MW; 0ACD7EF17540B8E2 CRC64;
 Query Match 60.7%; Score 34; DB 1; Length 426;
 Best Local Similarity 44.4%; Pred. No. 30;
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMW 9
 DQVVPAGLH 60
 DB 52 DQVVPAGLH 60
 RESULT 13
 ID YL15 VIBPA STANDARD; PRT; 441 AA.
 AC P46231.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 10-OCT-2003 (Rel. 42, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Hypothetical protein VP2115 (ORF3).
 SN VP2115.
 CS Vibrio parahaemolyticus.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIND 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Tagomori K.,
 RA Iijima Y., Nishima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae."

RL Lancet 361:743-749 (2003).
 RN [2]
 RP SEQUENCE OF 1-140 FROM N.A.
 RC STRAIN=BB22;
 RX MEDLINE=94292449; PubMed=8021208;
 RA McCarter L.L.;
 RT "MotY, a component of the sodium-type flagellar motor.";
 RL J. Bacteriol. 176:4219-4225 (1994).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: STRONG, TO H. INFLUENZAE HI0325.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AP005080; BAC60378.1; -.
 DR EMBL; U08949; AAA21571.1; -.
 DR InterPro; IPR004770; Antiporc_nhaC.
 DR InterPro; IPR001991; Na/diCO_symport.
 DR Pfam; PF03553; Na_H_antipporter; 1.
 DR PRINTS; PR00173; EDTNSPORT.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 195 215 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 SQ SEQUENCE 441 AA; 45961 MW; 451969FE307E4D46 CRC64;
 Query Match 58.9%; Score 33; DB 1; Length 441;
 Best Local Similarity 54.5%; Pred. No. 50;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMW 11
 DQVVPAGLH 60
 DB 52 DQVVPAGLH 60
 RESULT 14
 ID PHSL DESBA STANDARD; PRT; 513 AA.
 AC P13065;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Periplasmic [Nifese] hydrogenase large subunit (EC 1.12.99.6) (NifSe
 DE hydrogenase large chain).
 OS Desulfovibrio baculatus (Desulfomicrobium baculatus).
 CC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
 CC Desulfomicrobiaceae; Desulfomicrobium.
 CX NCBI_TaxID=899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88058744; PubMed=3316183;
 RA Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;
 RT "Cloning and sequencing of the genes encoding the large and small
 RT subunits of the periplasmic (Nifese) hydrogenase of Desulfovibrio
 RT baculatus.";
 RL J. Bacteriol. 169:5401-5407 (1987).
 RN [2]
 RP REVISIONS.
 RA Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;

RL J. Bacteriol. 170:4429-4429 (1988).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 RX MEDLINE=99306038; PubMed=10378275;
 RA Garcin E., Vernede X., Hatchikian E.C., Volbeda A., Frey M.,
 RA Pontecilla-Camps J.C.,
 RT Removal of the bridging ligand atom at the Ni-Fe active site of
 RT [NiFe] hydrogenase upon reduction with H₂, as revealed by X-ray
 RT structure analysis at 1.4-A resolution.";
 RL Structure 7:557-566 (1999).
 CC -1- CATALYTIC ACTIVITY: H(2) + acceptor = 2 H(+) + reduced acceptor.
 CC -1- COFACTOR: Nickel, 2 irons and selenocysteine. Iron 1 has three
 CC cyanide and carbon monoxide ligands. Iron 2 has three water
 CC ligands.
 CC -1- SUBUNIT: Heterodimer of a large and a small subunit.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- MISCELLANEOUS: PERHAPS THE LEADER OF THE SMALL SUBUNIT SERVES AS A
 CC TRANSPORT VEHICLE FOR BOTH SUBUNITS.
 CC -1- SIMILARITY: Belongs to the [NiFe]/[NiFeSe] hydrogenase large
 CC subunit family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M18271; AAA23375.2; -;
 DR PIR; A33101; HQDVLS.
 DR PDB; 1CC1; 01-JUN-99.
 DR InterPro; IPR001501; Ni_hdl.
 DR Pfam; PF00374; NiFeSe_Hases; 1.
 DR PROSITE; PS00507; NI_HENASE_L1; 1.
 DR PROSITE; PS00508; NI_HENASE_L2; 1.
 KW Oxidoreductase; Periplasmic; Metal-binding; Nickel; Iron; Selenium;
 KW Selenocysteine; 3D-structure.
 FT INIT MET 0
 FT METAL 51
 FT METAL 70
 FT METAL 70
 FT METAL 73
 FT METAL 73
 FT METAL 73
 FT METAL 444
 FT METAL 492
 FT METAL 495
 FT METAL 495
 FT METAL 495
 FT METAL 498
 FT SE_CYS 492
 FT STRAND 13
 FT STRAND 23
 FT STRAND 26
 FT TURN 34
 FT STRAND 36
 FT STRAND 46
 FT STRAND 50
 FT TURN 54
 FT TURN 59
 FT HELIX 62
 FT HELIX 66
 FT HELIX 74
 FT TURN 90
 FT TURN 90
 FT HELIX 95
 FT TURN 120
 FT HELIX 122
 FT TURN 125
 FT TURN 132
 FT TURN 139
 FT TURN 144
 FT HELIX 151
 FT STRAND 192
 FT TURN 194
 FT STRAND 196
 FT STRAND 197

FT HELIX 203
 FT TURN 222
 FT TURN 223
 FT HELIX 224
 FT HELIX 234
 FT STRAND 239
 FT STRAND 248
 FT STRAND 250
 FT TURN 253
 FT TURN 255
 FT STRAND 258
 FT STRAND 263
 FT STRAND 268
 FT TURN 272
 FT TURN 273
 FT STRAND 274
 FT STRAND 280
 FT STRAND 283
 FT STRAND 286
 FT TURN 288
 FT STRAND 290
 FT TURN 291
 FT STRAND 292
 FT STRAND 300
 FT STRAND 301
 FT HELIX 302
 FT STRAND 303
 FT STRAND 309
 FT TURN 311
 FT TURN 312
 FT TURN 314
 FT STRAND 315
 FT STRAND 318
 FT STRAND 322
 FT TURN 325
 FT TURN 327
 FT STRAND 328
 FT STRAND 331
 FT HELIX 334
 FT STRAND 344
 FT STRAND 347
 FT HELIX 357
 FT STRAND 363
 FT HELIX 364
 FT HELIX 371
 FT TURN 374
 FT TURN 401
 FT STRAND 406
 FT STRAND 415
 FT TURN 423
 FT TURN 424
 FT STRAND 426
 FT TURN 436
 FT STRAND 438
 FT TURN 447
 FT TURN 448
 FT TURN 451
 FT TURN 457
 FT HELIX 463
 FT TURN 468
 FT TURN 469
 FT STRAND 472
 FT TURN 476
 FT HELIX 480
 FT TURN 485
 FT STRAND 492
 FT HELIX 493
 FT STRAND 497
 SQ SEQUENCE 513 AA; 56683 MW; AC8285A6F80576FC CRC64;
 Query Match 58.9%; Score 33; DB 1; Length 513;
 Best Local Similarity 71.4%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 5 PGMHYS 11
 DB 297 PGLHYS 303
 RESULT 15
 MUTL BACSU
 ID MUTL BACSU STANDARD; PRT; 627 AA.
 AC P49850;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR BSU17050.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

NCBI_TaxID=1423;
[1]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis muts mutL operon: identification, nucleotide
sequence and mutagenesis";
RL Microbiology 142:2021-2029 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsche M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takamaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis".
RL Nature 390:249-256 (1997).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
DNA. It is required for dam-dependent methyl-directed DNA mismatch
repair. May act as a "molecular matchmaker", a protein that
promotes the formation of a stable complex between two or more
DNA-binding proteins in an ATP-dependent manner without itself
being part of a final effector complex (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL: U27343; AAB19236.1; -;
DR EMBL: Z99112; CAB13578.1; -;
DR PIR: A59663; A59663.
DR HSP: P23367; IBBN.
DR Subtilist; BGL1402; mutL.
DR HAMAP; MF_00149; -; 1.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_c1.
DR SMART; SW00387; HATPase_c1.
DR TIGRFAMs; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.

KW DNA repair; Complete proteome.
SQ SEQUENCE 627 AA; 70431 MW; 068A0509CC265343 CRC64;
Query Match 58.9%; Score 33; DB 1; Length 627;
Best Local Similarity 54.5%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 BEVVPXGMHYS 11
DB 488 EMIVPLTFHYS 498
Search completed: June 3, 2004, 11:49:53
Job time: 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds

(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-11

Perfect score: 56

Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_ricent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	73.2	413	11 Q8K289	Q8K289 mus musculus
2	39	69.6	1057	16 Q8CPJ4	Q8CPJ4 staphylococ
3	37	66.1	308	16 Q9X2E2	Q9X2E2 thermotoga
4	37	66.1	322	17 Q9HLH8	Q9HLH8 thermoplasma
5	37	66.1	1044	16 Q8DIH0	Q8DIH0 synchococ
6	36	64.3	208	2 Q46486	Q46486 corynebacte
7	36	64.3	252	16 Q8YWP1	Q8YWP1 anabaena sp
8	36	64.3	819	10 Q9AVK4	Q9AVK4 pium sativ
9	35	62.5	139	2 Q57489	Q57489 bacteroides
10	35	62.5	156	16 Q9PC35	Q9PC35 xyliella fas
11	35	62.5	156	16 Q87D36	Q87D36 xyliella fas
12	35	62.5	233	10 Q40479	Q40479 nicotiana t
13	35	62.5	237	10 Q9LW50	Q9LW50 nicotiana s
14	35	62.5	262	16 Q7WNB7	Q7WNB7 bordetella
15	35	62.5	262	16 Q7W023	Q7W023 bordetella
16	35	62.5	317	9 Q38317	Q38317 lactobacill

Q88G80 pseudomonas
Q9XVK4 caenorhabdi
Q9SA71 arabidopsis
Q8T207 methanopyru
Q7T1G5 gallus gall
Q9XN9 deinococcus
Q98HU6 rhizobium 1
Q9W92 deinococcus
Q9XAM3 streptomyce
Q27679 methanobact
Q9KBA1 bacillus ha
Q8CAL1 mus musculu
Q16912 caenorhabdi
Q876Z9 aspergillus
Q9Y0Y6 drosophila
Q9ELX6 cercopitheci
Q9YR01 ranid herpe
Q8BNL0 mus musculu
Q8YJ11 bruceella me
Q8G6X2 mus musculu
Q9V0L7 drosophila
Q12479 saccharomyc
Q8P6P8 xanthomonas
Q8KTQ4 candidatus
Q9RDC1 streptomyce
Q7V6Q4 prochloroco
Q8P1B0 xanthomonas
Q99011 prototheca
Q96N44 homo sapien

35 62.5 369 16 Q88G80
35 62.5 425 5 Q9XVK4
35 510 10 Q9SA71
35 62.5 2042 17 Q8T207
34 60.7 47 13 Q7T1G5
34 60.7 279 16 Q9XN9
34 60.7 285 16 Q98HU6
34 60.7 350 16 Q9W92
34 60.7 355 16 Q9XAM3
34 60.7 360 17 Q27679
34 60.7 425 16 Q9KBA1
34 60.7 484 11 Q8CAL1
34 60.7 495 5 Q16912
34 60.7 600 3 Q876Z9
34 60.7 637 5 Q9Y0Y6
34 60.7 678 12 Q9ELX6
34 60.7 748 12 Q9YR01
34 60.7 992 11 Q8BNL0
34 60.7 1028 16 Q8YJ11
34 60.7 1070 11 Q8G6X2
34 60.7 2438 5 Q9V0L7
33 58.9 156 3 Q12479
33 58.9 158 16 Q8P6P8
33 58.9 208 2 Q8KTQ4
33 58.9 216 16 Q9RDC1
33 58.9 245 16 Q7V6Q4
33 58.9 254 16 Q8P1B0
33 58.9 257 8 Q99011
33 58.9 273 4 Q96N44

ALIGNMENTS

RESULT 1

Q8K289 PRELIMINARY; PRT; 413 AA.

AC Q8K289; 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to expressed sequence A1987856.
GN TADA3L OR 111004B19RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC032195; AAH32195.1; -

DR MGD; MGI:1915724; Tada3l.

DR GO; GO:0030374; F:ligand-dependent nuclear receptor transcrip. .; IDA.

DR GO; GO:000515; F:protein binding; IPI.

SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 73.28; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 7.1;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11

||| ||| |||

Db 168 BEIFPLGKHYS 178

RESULT 2

Q8CPJ4 PRELIMINARY; PRT; 1057 AA.

AC Q8CPJ4; 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Carbamoyl-phosphate synthase large chain.
GN SE0879
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016746; AAC0476.1; -
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0005526; F:arginine biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005473; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS like.
DR InterPro; IPR000169; SHProt_acsite.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PRO0098; CPASE.
DR TIGRFAMs; TIGR01369; CPaseell_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Complete proteome.
SQ SEQUENCE 1057 AA; 117391 MW; 89447D8DB1CAE59 CRC64;

Query Match 59.6%; Score 39; DB 15; Length 1057;
Best Local Similarity 63.6%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHYS 11
Db 189 KEVVSNGLIHS 199
|||||:||||

RESULT 3
Q9X2E2 PRELIMINARY; PRT; 308 AA.
AC Q9X2E2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE FTSH protease activity modulator HFLK.
GN TM1822.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=23336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RA MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";

RL Nature 399:323-329 (1999).
DR EMBL; AE001819; AAD36885.1; -
DR PIR; A72207; A72207.
DR TIGR; TM1822; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR001972; StomatIn.
DR Pfam; PF01145; Band_7; 1.
DR PRINTS; PRO0721; STOMATIN.
DR SMART; SM00244; PHB; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 308 AA; 34778 MW; ADE03603E5101A9D CRC64;

Query Match 66.1%; Score 37; DB 16; Length 308;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMHY 10
Db 41 VVPSGIHY 48
|||||:||||

RESULT 4
Q9HLH8 PRELIMINARY; PRT; 322 AA.
ID Q9HLH8
AC Q9HLH8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucose-fructose oxidoreductase related protein.
GN TA0250.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum.";
RL Nature 407:508-513 (2000).
DR EMBL; AL45063; CAC11395.1; -
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;

Query Match 66.1%; Score 37; DB 17; Length 322;
Best Local Similarity 75.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMHY 10
Db 66 VVPDGLHY 73
|||||:||||

RESULT 5
Q8DIH0 PRELIMINARY; PRT; 1044 AA.
ID Q8DIH0
AC Q8DIH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.

```

OX NCBI_TaxID=32046;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=BP-1;
RC MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AF005374; BAC09170.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005215; F: transporter activity; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR001036; Acrlvin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00915; 2AG602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

Query Match 66.1%; Score 37; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
   |||:|:|:|
DB 843 BEVLPNGIGYS 853

RESULT 6
Q46486 PRELIMINARY; PRT; 208 AA.
AC Q46486;
CT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (Gora).
GN GCRA.
OS Corynebacterium xerosis, and
OC Corynebacterium striatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1725, 43770;
RN [1] SEQUENCE FROM N.A.
RP SPECIES=C.xerosis; STRAIN=M82B;
RC MEDLINE=96117603; PubMed=8559800;
RA Tauch A., Kassing F., Kalinowski J., Puhler A.;
RT "The Corynebacterium xerosis composite transposon Tn5432 consists of
RT two identical insertion sequences, designated IS1249, flanking the
RT erythromycin resistance gene ermCX."
RL Plasmid 34:119-131(1995).
RN [2] SEQUENCE FROM N.A.
RP SPECIES=C.striatum; STRAIN=M82B;
RC MEDLINE=20194806; PubMed=10732669;
RA Tauch A., Krieff S., Kalinowski J., Puhler A.;
RT "The 51,409-bp R-plasmid pTF10 from the multiresistant clinical
RT isolate Corynebacterium striatum M82B is composed of DNA segments
RT initially identified in soil bacteria and in plant, animal, and human
RT pathogens."
RL Mol. Gen. Genet. 263:1-11(2000).
DR EMBL; U21300; AAC95478.1; -.
DR EMBL; AF024666; CAG03390.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein, Plasmid.
SQ SEQUENCE 208 AA; 23012 MW; F1504BE1BCD8B5A6 CRC64;

Query Match 64.3%; Score 36; DB 2; Length 208;

```

```

Best Local Similarity 50.0%; Pred. No. 36;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 BEVVPXGMHYS 11
   ||:|:|:|:|
DB 130 DVIPEGRHYA 139

RESULT 7
Q8YWP1 PRELIMINARY; PRT; 252 AA.
ID Q8YWP1
AC Q8YWP1;
CT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein Alr1563.
GN Alr1563.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1] SEQUENCE FROM N.A.
RP MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003586; BAB77929.1; -.
DR PIR; AE2001; AE2001.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 252 AA; 28831 MW; 925572DA5D1CA519 CRC64;

Query Match 64.3%; Score 36; DB 16; Length 252;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
   ||:|:|:|:|
DB 235 EMIVPAGLHF 244

RESULT 8
Q9AVK4 PRELIMINARY; PRT; 819 AA.
ID Q9AVK4
AC Q9AVK4;
CT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE SCARECROW.
GN PSSCR.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=cv. Alaska;
RC MEDLINE=21231727; PubMed=11333309;
RA Sassa N., Matsushita Y., Nakamura T., Nyunoya H.;
RT "The Molecular Characterization and in situ Expression Pattern of Pea
RT SCARECROW Gene."
RL Plant Cell Physiol. 42:385-394(2001).
DR EMBL; AB048713; BAB39155.1; -.
DR TRANSFAC; T05513; -.
DR GO; GO:0009288; C: flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F: motor activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0001539; P: ciliary/flagellar motility; IEA.

```

```

DR InterPro; IPR001444; Flag_bb_rod.
DR InterPro; IPR005202; GRAS.
DR Pfam; PF03514; GRAS; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
SQ SEQUENCE 819 AA; 90372 MW; 41B67B6DC72ADFA CRC64;

Query Match          64.3%; Score 36; DB 10; Length 819;
Best Local Similarity 45.5%; Pred.No. 1.5e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
   :|||:||
DB 343 DDVVPTSLHFS 353

RESULT 9
Q57489 PRELIMINARY; PRT; 139 AA.
AC Q57489;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE DNA ligase (Fragment).
OS Bacteroides nodosus (Dichelobacter nodosus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Cardiobacteriales;
OC Cardiobacteriaceae; Dichelobacter.
NCBI_TaxID=870;
RN [1]
RP SEQUENCE FROM N.A.
RM MEDLINE=96020672; PubMed=7476204;
RA Moses E.K., Good R.T., Sinistaj M., Billington S.J., Langford C.J.,
RA Rood J.I.;
RT "A multiple site-specific DNA-inversion model for the control of Omp1
RT phase and antigenic variation in Dichelobacter nodosus.";
RL Mol. Microbiol. 17:183-196(1995).
RN [2]
RP SEQUENCE FROM N.A.
RM MEDLINE=96257263; PubMed=8654969;
RA Billington S.J., Sinistaj M., Cheetham B.F., Ayres A., Moses E.K.,
RA Katz M.E., Rood J.I.;
RT "Identification of a native Dichelobacter nodosus plasmid and
RT implications for the evolution of the vap regions.";
RL Gene 172:111-116(1996).
DR EMBL; U02462; AAB12366.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 1.
DR SMART; SM00292; BRCT; 1.
DR PROSITE; PS50172; BRCT; 1.
KW Ligase.
FT NON_TER
SQ SEQUENCE 139 AA; 15052 MW; E0E110AA4B7D4708 CRC64;

Query Match          62.5%; Score 35; DB 2; Length 139;
Best Local Similarity 55.6%; Pred.No. 37;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHYS 11
   :|||:||
DB 21 IVFAGVHWS 29

RESULT 10
Q9PC35 PRELIMINARY; PRT; 156 AA.
AC Q9PC35;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein Xf1950.
GN Xf1950.
OR Xvella fastidios.

```

Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xylella.
NCBI_TaxID=2371;
[1]
SEQUENCE FROM N.A.
RC STRAIN=9a5c;
MEDLINE=203165717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Canargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Faciniano A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Lemos E.G.M., Lenos M.V.P., Lopes S.A., Lopes C.R., Machado J.A.,
RA Lemos E.G.M., Lenos M.V.P., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Martins E.A.B.N., Madeira H.M.F., Matsumura A.Y.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsumura A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V. de Rosa A.J.M.,
RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Tsai S.M., Teuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000)
DR EMBL; AE004014; AF84752.1; -
DR PIR; D82618; D82618.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002545; ChEM.
DR PROSITE; PS50851; CHEM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 156 AA; 1714 MW; D8358619C6671A5D CRC64;

Query Match 62.5%; Score 35; DB 16; Length 156;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 REVVPXGNH 9
|||:|:|:
Db 119 REILPQGVH 127

RESULT 11
Q87D36 PRELIMINARY; PRT; 156 AA.
ID Q87D36; AC Q87D36;
DC Q87D36; DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-OCT-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
DE P08050.
OS Xylella fastidiosa (strain Temecual / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OC NCBI_TaxID=183190;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22421331; PubMed=12533478;
RX Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Canargo L.E.A., da Silva A.C.R., Moon D.H.,

[illegible]

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640440; CAE31621.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28907 MW; B3CA29331CB776B2 CRC64;

Query Match 62.5%; Score 35; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
Db 182 QEVVPYGOH 190

RESULT 15
Q7W0Z3 PRELIMINARY; PRT; 262 AA.
AC Q7W0Z3;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BPP0913.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Adkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640425; CAE40321.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28888 MW; 31CA2935EEB776BD CRC64;

Query Match 62.5%; Score 35; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9
Db 182 QEVVPYGOH 190

Search completed: June 3, 2004, 11:57:33
Job time : 30.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EGVVPXGMDYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseqp29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	52	96.3	11	5 ABB80561	Abb80561 Hepatitis
2	52	96.3	11	5 ABB80524	Abb80524 Hepatitis
3	52	96.3	11	5 ABB80529	Abb80529 Hepatitis
4	52	96.3	11	5 ABB80528	Abb80528 Hepatitis
5	52	96.3	11	5 ABB80562	Abb80562 Hepatitis
6	47	87.0	11	5 ABB80542	Abb80542 Hepatitis
7	47	87.0	11	5 ABB80543	Abb80543 Hepatitis
8	47	87.0	11	5 ABB80538	Abb80538 Hepatitis
9	46	85.2	11	5 ABB80525	Abb80525 Hepatitis
10	46	85.2	11	5 ABB80548	Abb80548 Hepatitis
11	46	85.2	11	5 ABB80521	Abb80521 Hepatitis
12	46	85.2	11	5 ABB80522	Abb80522 Hepatitis
13	46	85.2	11	5 ABB80547	Abb80547 Hepatitis
14	46	85.2	11	5 ABB80566	Abb80566 Hepatitis
15	46	85.2	11	5 ABB80556	Abb80556 Hepatitis
16	46	85.2	11	5 ABB80557	Abb80557 Hepatitis
17	46	85.2	11	5 ABB80551	Abb80551 Hepatitis
18	46	85.2	11	5 ABB80563	Abb80563 Hepatitis
19	46	85.2	11	5 ABB80565	Abb80565 Hepatitis
20	46	85.2	11	5 ABB80567	Abb80567 Hepatitis
21	46	85.2	11	5 ABB80559	Abb80559 Hepatitis
22	46	85.2	11	5 ABB80526	Abb80526 Hepatitis
23	46	85.2	11	5 ABB80564	Abb80564 Hepatitis
24	46	85.2	11	5 ABB80568	Abb80568 Hepatitis
25	45	83.3	11	5 ABB80523	Abb80523 Hepatitis

26 45 83.3 11 5 ABB80558 Hepatitis
27 45 83.3 11 5 ABB80560 Hepatitis
28 45 83.3 11 5 ABB80527 Hepatitis
29 44 81.5 11 5 ABB80534 Hepatitis
30 44 81.5 11 5 ABB80533 Hepatitis
31 41 75.9 11 5 ABB80536 Hepatitis
32 41 75.9 11 5 ABB80535 Hepatitis
33 41 75.9 11 5 ABB80540 Hepatitis
34 41 75.9 11 5 ABB80539 Hepatitis
35 40 74.1 11 5 ABB80549 Hepatitis
36 40 74.1 11 5 ABB80544 Hepatitis
37 40 74.1 11 5 ABB80537 Hepatitis
38 40 74.1 11 5 ABB80541 Hepatitis
39 40 74.1 11 5 ABB80553 Hepatitis
40 40 74.1 11 5 ABB80552 Hepatitis
41 40 74.1 11 5 ABB80545 Hepatitis
42 39 72.2 11 5 ABB80546 Hepatitis
43 39 72.2 11 5 ABB80554 Hepatitis
44 39 72.2 11 5 ABB80550 Hepatitis
45 39 72.2 11 5 ABB80555 Hepatitis

ALIGNMENTS

RESULT 1

ABB80561
ID ABB80561 standard; peptide; 11 AA.
AC ABB80561;
XX
DT 08-OCT-2002 (first entry)
XX
DB Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX
OS Synthetic.
FH Key
FT Modified-site 1 Location/Qualifiers
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 8 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX
SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
DB 1 EVVVPXGMDYS 11

RESULT 2
ID ABB80524 standard; peptide; 11 AA.
XX
AC ABB80524;
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX
SQ

SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
DB 1 EVVVPXGMDYS 11

RESULT 3
ID ABB80529 standard; peptide; 11 AA.
XX
AC ABB80529;
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX
SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 EEVVPXGMDYS 11
 |||||
 2b 1 EEVVPXGMDYS 11

RESULT 4

ABB80528
 ID ABB80528 standard; peptide; 11 AA.

AC
 KC ABB80528;

2T 08-OCT-2002 (first entry)

2E Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

2X Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

2S Synthetic.

2H Key Location/Qualifiers

2T Modified-site 1 /note= "N-terminal acetyl"

2T Modified-site 5 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

2T Misc-difference 8 /note= "D-form residue"

2T Modified-site 11 /note= "C-terminal amide"

2X WO200208251-A2.

2X 31-JAN-2002.

2X 19-JUL-2001; 2001WO-US023169.

2X 21-JUL-2000; 2000US-0220101P.

2X (CORV-) CORVAS INT INC.

2X Lim-Wilby M, Levy OE, Brunck TK;

2X WPI; 2002-361643/39.

2X Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

2X Claim 17; Page 64; 69pp; English.

2X The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

2X Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.0024;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 EEVVPXGMDYS 11

2b 1 EEVVPXGMDYS 11

RESULT 5

ABB80562

ID ABB80562 standard; peptide; 11 AA.

XX ABB80562;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site 1 /note= "N-terminal acetyl"

XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

XX Modified-site 8 /note= "Oxymethionine"

XX Misc-difference 8 /note= "D-form residue"

XX Misc-difference 9 /note= "D-form residue"

XX Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

XX Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.0024;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 EEVVPXGMDYS 11

2b 1 EEVVPXGMDYS 11

RESULT 6

ABB80542

ID ABB80542 standard; peptide; 11 AA.

XX ABB80542;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
 DE
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 DS Synthetic.
 XX
 FH Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT virucide."
 FT
 FT Misc-difference 8 /note= "D-form residue"
 FT
 FT Misc-difference 9 /note= "D-form residue"
 FT
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. NO. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGQDYS 11
 RESULT 7
 ABB80543
 ID ABB80543 standard; peptide; 11 AA.
 AC ABB80543;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX

FH Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT
 FT Misc-difference 8 /note= "D-form residue"
 FT
 FT Misc-difference 9 /note= "D-form residue"
 FT
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. NO. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGQDYS 11
 RESULT 8
 ABB80538
 ID ABB80538 standard; peptide; 11 AA.
 AC ABB80538;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT

T	Misc-difference	9
T	/note= "D-form residue"	
T	Modified-site	11
T	/note= "C-terminal amide"	
N	WO200208251-A2.	
X		
D	31-JAN-2002.	
X		
F	19-JUL-2001; 2001WO-US23169.	
X		
R	21-JUL-2000; 2000US-0220101P.	
X	(CORV-) CORVAS INT INC.	
A		
X	Lim-Wilby M, Levy OE, Brunc TK;	
I	WPI; 2002-361643/39.	
R		
X	Novel peptide compound having hepatitis C virus protease inhibitory	
T	activity useful for treating disorders associated with hepatitis C virus	
T	protease.	
T		
S	Claim 17; Page 64; 69pp; English.	
X		
C	The sequence represents a peptide compound of the invention having	
C	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the	
C	invention are alpha-ketoamide peptide analogues. The peptides have	
C	virucide activity, and are useful for treating and in the manufacture of	
C	a medicament to treat disorders associated with HCV protease. A	
C	pharmaceutical composition comprising the peptide as an active ingredient	
X	is useful for treating disorders associated with hepatitis C virus	
X		
Q	Sequence 11 AA;	
	Query Match	87.0%; Score 47; DB 5; Length 11;
	Best Local Similarity	90.9%; Pred.No. 0.023;
	Matches	10; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Y	1 EEVPXGMDYS 11	
b	1 EEVPXGQDYS 11	
RESULT	9	
B	B80525	
D	ABB80525 standard; peptide; 11 AA.	
X		
C	ABB80525;	
X		
T	08-OCT-2002 (first entry)	
X		
E	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.	
X		
W	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;	
W	virucide.	
X		
S	Synthetic.	
X		
K	Key	Location/Qualifiers
T	Modified-site	1/note= "N-terminal acetyl"
T		
T	Modified-site	6/note= "Norvalyl carbonyl forming keto-amide linkage with
T		residue 7"
T	Misc-difference	8
T	/note= "D-form residue"	
T	Modified-site	11
T	/note= "C-terminal amide"	
X		
N	WO200208251-A2.	
X		
D	31-JAN-2002.	

XX	19-JUL-2001; 2001WO-US023169.
PF	21-JUL-2000; 2000US-0220101P.
XX	(CORV-) CORVAS INT INC.
XX	Lam-Wilby M, Levy OE, Brunck TK;
XX	WPI; 2002-361643/39.
XX	Novel peptide compound having hepatitis C virus protease inhibitory
PT	activity useful for treating disorders associated with hepatitis C virus
PT	protease.
XX	Claim 17; Page 64; 69pp; English.
PS	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a medicament to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus
XX	
XX	Sequence 11 AA;
QY	Query Match 85.2%; Score 46; DB 5; Length 11;
DB	Best Local Similarity 90.9%; Pred. No. 0.036; 1; Indels 0; Gaps 0;
	Matches 10; Conservative 0; Mismatches 1;
QY	1 EEVVPXGMDYS 11
DB	1 EEVVPXGMSYS 11
RESULT 10	
ABB80548	
ID	ABB80548 standard; peptide; 11 AA.
AC	ABB80548;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
XX	
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW	virucide.
XX	
OS	Synthetic.
XX	
Key	Location/Qualifiers
FT	Modified-site 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Misc-difference 9 /note= "D-form residue"
FT	Modified-site 11 /note= "C-terminal amide"
FT	
XX	
PN	W0200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US023169.
XX	
PR	21-JUL-2000; 2000US-0220101P.
XX	
EA	(CORV-) CORVAS INT INC.
XX	
PI	Lam-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.
DR
XX
PT Novel peptide compound having hepatitis C virus protease inhibitor
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
PS
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ
Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMDYS 11
RESULT 11
ABB80521 standard; peptide; 11 AA.
XX
XX ABB80521;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitor
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ
Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMDYS 11
RESULT 12
ABB80522 standard; peptide; 11 AA.
XX
XX ABB80522;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8
FT /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX
XX PD 31-JAN-2002.
XX
XX PF 19-JUL-2001; 2001WO-US023169.
XX
XX PR 21-JUL-2000; 2000US-0220101P.
XX
XX PA (CORV-) CORVAS INT INC.
XX
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX
XX DR WPI; 2002-361643/39.
XX
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX PS Claim 17; Page 65; 69pp; English.
XX
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPKGMDS 11
| | | | | | | | | |
Db 1 EEVVPKGSYS 11

Search completed: June 3, 2004, 11:48:23
Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/2/iaa/5A COMB.pdp.*
- 2: /cgm2_6/ptodata/2/iaa/5B COMB.pdp.*
- 3: /cgm2_6/ptodata/2/iaa/6A COMB.pdp.*
- 4: /cgm2_6/ptodata/2/iaa/6B COMB.pdp.*
- 5: /cgm2_6/ptodata/2/iaa/PCRU COMB.pdp.*
- 6: /cgm2_6/ptodata/2/iaa/backfileseq.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	66.7	181	4	US-09-134-000C-4848
2	36	66.7	341	3	US-08-853-948B-4
3	36	66.7	348	3	US-08-853-948B-5
4	36	66.7	1049	4	US-09-697-367-24
5	36	66.7	1049	4	US-09-394-272-10
6	36	66.7	1068	2	US-08-429-054A-11
7	36	66.7	1068	2	US-08-718-777-7
8	36	66.7	1068	3	US-09-051-341-7
9	36	66.7	1068	4	US-09-394-272-8
10	36	66.7	1081	4	US-09-394-272-4
11	36	66.7	1083	4	US-09-394-272-11
12	36	66.7	1084	4	US-09-394-272-9
13	34	63.0	140	3	US-08-569-147-76
14	34	63.0	140	3	US-08-569-147-82
15	34	63.0	1065	4	US-09-252-991A-31637
16	34	63.0	3472	4	US-09-408-020-4
17	33	61.1	59	4	US-08-963-851-14
18	33	61.1	237	4	US-09-540-236-2677
19	33	61.1	303	4	US-09-134-000C-4318
20	33	61.1	378	1	US-08-070-165F-8
21	33	61.1	378	2	US-08-885-418-8
22	33	61.1	473	4	US-09-252-991A-26805
23	33	61.1	765	4	US-09-839-989-4
24	33	61.1	765	4	US-10-273-992-4
25	33	61.1	801	3	US-09-383-630-6
26	33	61.1	811	4	US-09-819-989-2
27	33	61.1	811	4	US-10-273-992-2

28	61.1	883	4	US-09-667-373-4	Sequence 4, Appli
29	59.3	65	6	517197-51	Patent No. 517197
30	59.3	102	2	US-08-580-988A-23	Sequence 23, Appli
31	59.3	152	2	US-08-460-694-4	Sequence 4, Appli
32	59.3	152	2	US-08-460-744-4	Sequence 4, Appli
33	59.3	152	3	US-07-667-711B-4	Sequence 4, Appli
34	59.3	173	1	US-08-193-977-7	Sequence 7, Appli
35	59.3	189	2	US-08-464-517-21	Sequence 21, Appli
36	59.3	189	2	US-08-246-361A-21	Sequence 21, Appli
37	59.3	189	3	US-08-463-772-21	Sequence 21, Appli
38	59.3	189	5	PCT-US93-05000-21	Sequence 21, Appli
39	59.3	231	3	US-08-926-842B-20	Sequence 20, Appli
40	59.3	236	2	US-08-464-517-22	Sequence 22, Appli
41	59.3	236	2	US-08-246-361A-22	Sequence 22, Appli
42	59.3	236	3	US-08-463-772-22	Sequence 22, Appli
43	59.3	236	5	PCT-US93-05000-22	Sequence 22, Appli
44	59.3	240	3	US-08-926-842B-21	Sequence 21, Appli
45	59.3	241	4	US-09-107-532A-4172	Sequence 4172, Ap

ALIGNMENTS

RESULT 1

US-09-134-000C-4848
; Sequence 4848, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1998-05-15
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6912
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4848
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 66.7%; Score 36; DB 4; Length 181;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDY 10
Db 145 EEVVPISDY 154

RESULT 2

US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAWA, Toyota
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURE:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; OTHER INFORMATION: Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe,

; OTHER INFORMATION: Pro, Ser, Thr, Trp, Tyr, or Val

US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|:
Db 228 VIPPGMDFS 236

RESULT 3

US-08-853-948B-5

; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|:
Db 234 VIPPGMDFS 242

RESULT 4

US-09-697-367-24

; Sequence 24, Application US/09697367
; Patent No. 623015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Cai, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Tarczynski, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BB1166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|:
Db 217 VIPPGMDFS 225

RESULT 5

US-09-394-272-10

; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|:
Db 436 VIPPGMDFS 444

RESULT 6

US-08-429-054A-11

; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Musserlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000

DB 435 VIPGMDPS 443

RESULT 8
US-09-051-341-7
; Sequence 7, Application US/09051341
; Patent No. 6124528
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, C. K.
; TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rae-Venter Law Group, P.C.
; STREET: 260 Sheridan Avenue, Suite 440
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,341
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17351
; FILING DATE: 25-OCT-1996
; APPLICATION NUMBER: US 08/549,016
; FILING DATE: 27-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/372,200
; FILING DATE: 12-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara Rae-Venter, Ph.D.,
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE.110.02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)328-4400
; TELEFAX: (415)328-4477
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-051-341-7

Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

RESULT 9
US-09-394-272-8
; Sequence 8, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14

DB 435 VIPGMDPS 443

QY 3 VVPXGMDS 11
|:|||||

US-08-718-777-7

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

RESULT 7
US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE
; TITLE OF INVENTION: PHOSPHATE
; TITLE OF INVENTION: SYNTHASE IN PLANTS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Barbara Rae-Venter
; STREET: 260 Sheridan Avenue, Suite 440
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,777
; FILING DATE: NOT YET ASSIGNED
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,471
; FILING DATE: 27-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara Rae-Venter
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE.072.02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)328-4400
; TELEFAX: (415)328-4477
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-777-7

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||

US-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDS 11
|:|||||
Db 435 VIPGMDPS 443

TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
JS-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 8
 LENGTH: 1068
 TYPE: PRT
 ORGANISM: Zea mays
 US-09-394-272-8

Query Match 66.7%; Score 36; DB 4; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 83;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:||||:
 Db 435 VIPPGMDFS 443

RESULT 10

US-09-394-272-4
 Sequence 4, Application US/09394272
 Patent No. 6472588

GENERAL INFORMATION:
 APPLICANT: Haigler, Candace H.
 APPLICANT: Holaday, A. Scott
 TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
 TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
 FILE REFERENCE: 201304/1000
 CURRENT APPLICATION NUMBER: US/09/394,272
 CURRENT FILING DATE: 1999-09-10
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4
 LENGTH: 1081
 TYPE: PRT
 ORGANISM: Craterostigma plantagineum
 US-09-394-272-4

Query Match 66.7%; Score 36; DB 4; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 83;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:||||:
 Db 445 VIPPGMDFS 453

RESULT 11

US-09-394-272-11
 Sequence 11, Application US/09394272
 Patent No. 6472588

GENERAL INFORMATION:
 APPLICANT: Haigler, Candace H.
 APPLICANT: Holaday, A. Scott
 TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
 TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
 FILE REFERENCE: 201304/1000
 CURRENT APPLICATION NUMBER: US/09/394,272
 CURRENT FILING DATE: 1999-09-10
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 11
 LENGTH: 1083
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana
 US-09-394-272-11

Query Match 66.7%; Score 36; DB 4; Length 1083;
 Best Local Similarity 66.7%; Pred. No. 83;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:||||:
 Db 483 VIPPGMDFS 491

RESULT 12

US-09-394-272-9
 Sequence 9, Application US/09394272
 Patent No. 6472588

GENERAL INFORMATION:
 APPLICANT: Haigler, Candace H.
 APPLICANT: Holaday, A. Scott
 TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
 TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
 FILE REFERENCE: 201304/1000
 CURRENT APPLICATION NUMBER: US/09/394,272
 CURRENT FILING DATE: 1999-09-10
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 9
 LENGTH: 1084
 TYPE: PRT
 ORGANISM: Oryza sativa
 US-09-394-272-9

Query Match 66.7%; Score 36; DB 4; Length 1084;
 Best Local Similarity 66.7%; Pred. No. 83;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:||||:
 Db 453 VIPPGMDFS 461

RESULT 13

US-08-569-147-76
 Sequence 76, Application US/08569147
 Patent No. 6180377

GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: HUMANISED ANTIBODIES
 NUMBER OF SEQUENCES: 95
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ADDRESSEE: No. 6180377ris, LLP
 STREET: One Liberty Place - 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: U.S.A.
 ZIP: 19103

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/569,147
 FILING DATE: 25-March-1996

CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Trujillo, Doreen Yanko
 REGISTRATION NUMBER: 35,719
 REFERENCE/DOCKET NUMBER: CARP-0047
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439 76:
 INFORMATION FOR SEQ ID NO: 76:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 140 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-569-147-76

Query Match 63.0%; Score 34; DB 3; Length 140;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31637
Query Match 63.0%; Score 34; DB 4; Length 1065;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMDYS 11
Db 324 PQGMDYS 330

Search completed: June 3, 2004, 12:03:08
Job time : 12.8 secs

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31637
Query Match 63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMDY 10
Db 122 VVPTGFDY 129

RESULT 14
JS-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377is, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
JS-08-569-147-82

Query Match 63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMDY 10
Db 122 VVPTGFDY 129

RESULT 15
US-09-252-991A-31637
; Sequence 31637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31637
; LENGTH: 1065
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-12
Perfect score: 54
Sequence: 1 BEVFXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 28138677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	96.3	11	12	US-09-909-164-8
2	52	96.3	11	12	US-09-909-164-12
3	52	96.3	11	12	US-09-909-164-13
4	47	87.0	11	12	US-09-909-164-22
5	47	87.0	11	12	US-09-909-164-26
6	47	87.0	11	12	US-09-909-164-27
7	46	85.2	11	12	US-09-909-164-5
8	46	85.2	11	12	US-09-909-164-6
9	46	85.2	11	12	US-09-909-164-9
10	46	85.2	11	12	US-09-909-164-10
11	46	85.2	11	12	US-09-909-164-31
12	46	85.2	11	12	US-09-909-164-32
13	46	85.2	11	12	US-09-909-164-35
14	46	85.2	11	12	US-09-909-164-40
15	46	85.2	11	12	US-09-909-164-41

16	46	85.2	11	12	US-09-909-164-45	Sequence 45, Appl
17	46	85.2	11	12	US-09-909-164-46	Sequence 46, Appl
18	46	85.2	11	12	US-09-909-164-47	Sequence 47, Appl
19	46	85.2	11	12	US-09-909-164-48	Sequence 48, Appl
20	46	85.2	11	12	US-09-909-164-49	Sequence 49, Appl
21	46	85.2	11	12	US-09-909-164-50	Sequence 50, Appl
22	46	85.2	11	12	US-09-909-164-51	Sequence 51, Appl
23	46	85.2	11	12	US-09-909-164-52	Sequence 52, Appl
24	45	83.3	11	12	US-09-909-164-7	Sequence 7, Appl
25	45	83.3	11	12	US-09-909-164-11	Sequence 11, Appl
26	44	81.5	11	12	US-09-909-164-17	Sequence 17, Appl
27	44	81.5	11	12	US-09-909-164-18	Sequence 18, Appl
28	41	75.9	11	12	US-09-909-164-19	Sequence 19, Appl
29	41	75.9	11	12	US-09-909-164-20	Sequence 20, Appl
30	41	75.9	11	12	US-09-909-164-23	Sequence 23, Appl
31	41	75.9	11	12	US-09-909-164-24	Sequence 24, Appl
32	40	74.1	11	12	US-09-909-164-21	Sequence 21, Appl
33	40	74.1	11	12	US-09-909-164-25	Sequence 25, Appl
34	40	74.1	11	12	US-09-909-164-28	Sequence 28, Appl
35	40	74.1	11	12	US-09-909-164-29	Sequence 29, Appl
36	40	74.1	11	12	US-09-909-164-33	Sequence 33, Appl
37	40	74.1	11	12	US-09-909-164-36	Sequence 36, Appl
38	40	74.1	11	12	US-09-909-164-37	Sequence 37, Appl
39	40	74.1	11	12	US-09-909-164-43	Sequence 43, Appl
40	39	72.2	11	12	US-09-909-164-30	Sequence 30, Appl
41	39	72.2	11	12	US-09-909-164-34	Sequence 34, Appl
42	39	72.2	11	12	US-09-909-164-38	Sequence 38, Appl
43	39	72.2	11	12	US-09-909-164-39	Sequence 39, Appl
44	39	72.2	11	12	US-09-909-164-42	Sequence 42, Appl
45	39	72.2	11	12	US-09-909-164-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-909-164-8
; Sequence 8, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)-(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (9)-(9)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)

; OTHER INFORMATION: AMIDATION
US-09-909-164-8

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMDYS 11

RESULT 2

US-09-909-164-12
; Sequence 12, Application US/09909164
; Publication No. US2002068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-12

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMDYS 11

RESULT 3

US-09-909-164-13
; Sequence 13, Application US/09909164
; Publication No. US2002068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US

; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-13

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMDYS 11

RESULT 4

US-09-909-164-22
; Sequence 22, Application US/09909164
; Publication No. US2002068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-12

Query Match 96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
| | | | | | | | | | |
Db 1 EEVVPXGMDYS 11

; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-22

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 5

US-09-909-164-26
; Sequence 26, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 26

Length: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-26

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 6

US-09-909-164-27
; Sequence 27, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita

; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 27

Length: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-27

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 7

US-09-909-164-5
; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5

Length: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

OTHER INFORMATION: 11-mer synthesized according to example 1

```
/ OTHER INFORMATION: norvaline-(CO)
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(11)
/ OTHER INFORMATION: AMIDATION
US-09-909-164-5
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMDYS 11
Db 1 EVVVPXGMSYS 11

RESULT 8
US-09-909-164-6
/ Sequence 6, Application US/09909164
/ Publication No. US20020068702A1
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
/ APPLICANT: Levy, Odile E
/ APPLICANT: Bruck, Terence K
/ TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
/ FILE REFERENCE: IN01192-US
/ CURRENT APPLICATION NUMBER: US/09/909,164
/ PRIOR FILING DATE: 2003-03-25
/ PRIOR FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 62
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 6
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: 11-mer synthesized according to example 1
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(11)
/ OTHER INFORMATION: ACETYLATION
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (11)..(11)
/ OTHER INFORMATION: AMIDATION
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (6)..(6)
/ OTHER INFORMATION: norvaline-(CO)
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (8)..(8)
/ OTHER INFORMATION: D-amino acid
/ US-09-909-164-9
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMDYS 11
Db 1 EVVVPXGMSYS 11

RESULT 10
US-09-909-164-10
/ Sequence 10, Application US/09909164
/ Publication No. US20020068702A1
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
/ APPLICANT: Levy, Odile E
/ APPLICANT: Bruck, Terence K
/ TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
/ FILE REFERENCE: IN01192-US
/ CURRENT APPLICATION NUMBER: US/09/909,164
/ PRIOR FILING DATE: 2003-03-25
/ PRIOR FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 62
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 10
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: 11-mer synthesized according to example 1
/ NAME/KEY: MOD RES
/ LOCATION: (1)..(11)
/ OTHER INFORMATION: ACETYLATION
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (11)..(11)
/ OTHER INFORMATION: AMIDATION
/ US-09-909-164-6
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMDYS 11
Db 1 EVVVPXGMSYS 11

RESULT 9
US-09-909-164-9
/ Sequence 9, Application US/09909164
/ Publication No. US20020068702A1
/ GENERAL INFORMATION:
/ APPLICANT: Corvas International, Inc.
/ APPLICANT: Lim-Wilby, Marguerita
```

```

; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
JS-09-909-164-10

Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
Db 1 BEVVPXGMSYS 11

RESULT 11
JS-09-909-164-31
; Sequence 31, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: 11-mer synthesized according to example 1
US-09-909-164-32

Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
Db 1 BEVVPXGTDYS 11

RESULT 13
US-09-909-164-35
; Sequence 35, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)

```

```
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-35
```

```
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 BEVVPXGMDYS 11
        |||||
Db      1 BEVVPXGSDYS 11
```

RESULT 14

```
US-09-909-164-40
; Sequence 40, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 40
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-40
```

```
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 BEVVPXGMDYS 11
        |||||
Db      1 BEVVPXGSDYS 11
```

RESULT 15

```
US-09-909-164-41
; Sequence 41, Application US/09909164
```

```
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 41
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-41
```

```
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 BEVVPXGMDYS 11
        |||||
Db      1 BEVVPXGSDYS 11
```

```
Search completed: June 3, 2004, 12:57:16
Job time : 34.7333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004; 11:35:47 ; Search time 9 seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EVVFXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	40	74.1	156	2	S54619	hypothetical prote
2	38	70.4	363	2	D69551	conserved hypothet
3	36	66.7	102	2	A42452	v1 protein - tobac
4	36	66.7	341	2	S72649	sucrose-phosphate
5	36	66.7	348	2	S72650	sucrose-phosphate
6	36	66.7	460	2	G96784	unknown protein P2
7	36	66.7	1049	2	JC4783	sucrose-phosphate
8	36	66.7	1068	1	QJ1329	sucrose-phosphate
9	36	66.7	1081	2	T09837	sucrose-phosphate
10	36	66.7	1083	2	T04062	sucrose-phosphate
11	36	66.7	1084	2	T04103	sucrose-phosphate
12	35	64.8	425	2	T24111	hypothetical prote
13	35	64.8	433	2	H87660	peptidoglycan-bind
14	35	64.8	440	2	H72784	probable alkaline
15	35	64.8	1150	2	T20173	hypothetical prote
16	35	64.8	1474	2	P69009	probable membrane
17	35	64.8	2747	2	B49132	fat facets (faf) s
18	34	63.0	99	2	S00210	plastocyanin b - L
19	34	63.0	155	2	S38258	plastocyanin b pre
20	34	63.0	168	2	S38268	plastocyanin b pre
21	34	63.0	290	2	A33104	6-O-methylguanine-D
22	34	63.0	290	2	D38182	hypothetical prote
23	34	63.0	296	2	T72745	probable hexosyltr
24	34	63.0	357	1	G69290	L-lactate dehydrog
25	34	63.0	366	2	G59350	ABC transporter At
26	34	63.0	565	2	E86665	phenylalanine-tRNA
27	34	63.0	586	2	A70164	succinate dehydrog
28	34	63.0	587	2	F81138	disease resistance
29	34	63.0	906	2	T48898	

ALIGNMENTS

RESULT 1

S54619

hypothetical protein YOR013w - Yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3

C/Species: Saccharomyces cerevisiae

C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C/Accession: S54619; S66879

R/de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A/Reference number: S54619

A/Accession: S54619

A/Molecule type: DNA

A/Residues: 1-156 <DEH>

A/Cross-references: EMBL:X87331; NID:g1041652; PIDN:CAA60762.1; PID:g829123

R/de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A/Reference number: S66877

A/Accession: S66879

A/Molecule type: DNA

A/Residues: 1-156 <DEW>

A/Cross-references: EMBL:Z74920; NID:g1420109; PIDN:CAA9201.1; PID:g1420111; MIPS:YOR01

A/Experimental source: strain S288C

C/Genetics:

A/Cross-references: SGD:S0005539

A/Map position: 15R

C/Superfamily: hypothetical protein YOR013w

Query Match

Best Local Similarity 74.1%; Score 40; DB 2; Length 156;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDY 10

Db 50 EVVFXGMDY 58

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

```

A:Molecule type: DNA
A:Residues: 1-363 <XLE>
A/Cross-references: GB:AE001109; GB:AE000782; NID:g2689432; PIDN:AB91255.1; PID:g265068

Query Match      70.4%; Score 38; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 6.8;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
    | : || | : ||
Db 120 ENIVPYGIDFS 130

RESULT 3
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C/Species: tobacco yellow dwarf virus
C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C/Accession: A42452
R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A/Reference number: A42452; MUID:92188538; PMID:1546458
A/Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A/Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match      66.7%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 4.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMDYS 11
    | : || | : ||
Db 7 QVPSGGINYS 16

RESULT 4
S72649
sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Citrus unshiu (fragment)
C/Species: Citrus unshiu
C/Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C/Accession: S72649
R/Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
Mol. Gen. Genet. 252, 346-351, 1996
A/Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
A/Reference number: S72648; MUID:96439842; PMID:8842155
A/Accession: S72649
A/Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-341 <KOM>
A/Cross-references: EMBL:AB006319; NID:g2588891; PIDN:BA23215.1; PID:g2588892
A/Experimental source: fruit, cv. Miyagawa-wase
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
C/Genetics:
A:Gene: SPS2
C/Function:
A/Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A/Pathway: sucrose biosynthesis
C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F/1-341/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SPS>

Query Match      66.7%; Score 36; DB 2; Length 341;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
    | : || | : ||
Db 228 VIPPGMDFS 236

RESULT 5

```

```

;Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
;Accession: J04783
;Valdez-Alarcon, J.J.; Ferrando, M.; Salerno, G.; Jimenez-Moraila, B.; Herreeta-Estrella
;ene 170, 217-222, 1996
;Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
;Reference number: J04783; MUID:96235138; PMID:8666248
;Accession: J04783
;Molecule type: mRNA
;Residues: 1-1049 <VAL>
;Cross-references: GB:U33175; NID:G1449931; PIDN:AA049379.1; PID:G988270
;Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrosephosphate-UDPglucosyltr
;Comment: This enzyme catalyzes the formation of sucrose-phosphate form UDP-glucose and
;Genetics:
;Gene: Spel
;Introns: 24/1; 103/3; 183/3; 205/3; 435/3; 475/3; 519/3; 596/3; 617/3; 931/3; 9
;Function:
;Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
;Pathway: sucrose biosynthesis
;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
;Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
;180-663/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

> 3 VVPXGMDYS 11
|:|:|:|:|
>b 436 VIPFGMDFS 444

RESULT 8
J01329
sucrose-phosphate synthase (EC 2.4.1.14) - maize
;Species: Zea mays (maize)
;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
;Accession: J01329; PQ0260
;Worrell, A.C.; Bruneau, J.M.; Summerfelt, K.; Boersig, M.; Voelker, T.A.
;Plant Cell 3, 1121-1130, 1991
;Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
;Reference number: J01329; MUID:92338837; PMID:1840396
;Accession: J01329
;Molecule type: mRNA
;Residues: 1-1068 <WOR>
;Cross-references: GB:M97550; NID:G168625; PIDN:AAA33513.1; PID:G168626
;Accession: PQ0260
;Molecule type: protein
;Residues: 71-74;206-212;471-481;872-892 <WOR>
;Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
;Function:
;Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
;Pathway: sucrose biosynthesis
;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
;Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
;178-666/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

> 3 VVPXGMDYS 11
|:|:|:|:|
>b 435 VIPFGMDFS 443

RESULT 9
J09837
sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantagineum
;Species: Craterostigma plantagineum
;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
;Accession: T09837
;Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

```

```

Plant Physiol. 115, 113-121, 1997
;Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to suga
;Reference number: Z15874; MUID:97451773; PMID:9306694
;Accession: T09837
;Status: preliminary; translated from GB/EMBL/DBDJ
;Molecule type: mRNA
;Residues: 1-1081 <ING>
;Cross-references: EMBL:Y11795; NID:G2190349; PIDN:CAA72491.1; PID:G2190350
;Experimental source: ABA-treated callus
;Genetics:
;Gene: spe2
;Function:
;Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
;Pathway: sucrose biosynthesis
;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
;Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
;176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSS>

Query Match 66.7%; Score 36; DB 2; Length 1081;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

> 3 VVPXGMDYS 11
|:|:|:|:|
>b 445 VIPFGMDFS 453

RESULT 10
T04062
sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
;Accession: T04062
;Bavan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
;submitted to the Protein Sequence Database, March 1999
;Reference number: Z15184
;Accession: T04062
;Molecule type: DNA
;Residues: 1-1083 <BEV>
;Cross-references: EMBL:AL049487
;Experimental source: cultivar Columbia; BAC clone F28M11
;Genetics:
;Map position: 4
;Introns: 86/3; 116/3; 255/3; 322/2; 482/3; 526/3; 570/3; 629/3; 647/3; 658/3; 949/3; 9
;Note: F28M11.40
;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
;230-714/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

> 3 VVPXGMDYS 11
|:|:|:|:|
>b 483 VIPFGMDFS 491

RESULT 11
T04103
sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
;Species: Oryza sativa (rice)
;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
;Accession: T04103
;Sakamoto, M.; Satorawa, T.; Kishimoto, N.; Higo, K.; Shinada, H.; Fujimura, T.
;Plant Sci. 112, 207-217, 1995
;Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene that
;Reference number: Z15212
;Accession: T04103
;Status: preliminary; translated from GB/EMBL/DBDJ
;Molecule type: DNA
;Residues: 1-1084 <SAK>
;Cross-references: EMBL:D45890; PIDN:BAA08304.1
;Experimental source: subsp. Japonica

```


C:Genetics:
A:Gene: Sps1

A:Map position: 1
A:Introns: 120/3; 200/2; 221/3; 452/3; 492/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase, hexosyltransferase
F:196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1084;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMDYS 11
|:|:|:|:|:|:|
Db 453 VVPGGMDYS 461

RESULT 12

T24111 hypothetical protein R10D12.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19842
A:Accession: T24111

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-425 <WIL>

A:Cross-references: EMBL:Z81109; PIDN:CA803241.1; GSPDB:GN00023; CESP:R10D12.10

A:Experimental source: clone R10D12

C:Genetics:

A:Gene: CESP:R10D12.10

A:Map position: 5

A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match

Best Local Similarity 64.8%; Score 35; DB 2; Length 425;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVFXGMDY 10
|:|:|:|:|:|:|
Db 335 EQVPGGLQY 344

RESULT 13

H87660

Peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: H87660

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87429; MUID:z1173698; PMID:11259647

A:Accession: H87660

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <STO>

A:Cross-references: GB:AE005673; NID:gl3425020; PIDN:AAK25284.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3322

Query Match

Best Local Similarity 64.8%; Score 35; DB 2; Length 433;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVFXGMDYS 11
|:|:|:|:|:|:|
Db 265 EVLPFGFDYS 276

RESULT 14

H72784

Probable alkaline proteinase APO0263 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: H72784

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: H72784

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <KAW>

A:Cross-references: DDBJ:AP000058; NID:gs103388; PIDN:BAA79178.1; PID:gs103657

A:Experimental source: strain K1

C:Genetics:

A:Gene: APO0263

C:Superfamily: subtilisin; subtilisin homology

Query Match 64.8%; Score 35; DB 2; Length 440;

Best Local Similarity 66.7%; Pred. No. 35;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDY 10
|:|:|:|:|:|:|
Db 120 EVLPWGVDY 128

RESULT 15

T20173

hypothetical protein CS3A5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T20173; T23857

R:Northmore, B.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19232

A:Accession: T20173

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1150 <WIL>

A:Cross-references: EMBL:Z81486; PIDN:CA803994.1; GSPDB:GN00023; CESP:CS3A5.2

A:Experimental source: clone CS3A5

R:Matthews, L.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z19808

A:Accession: T23857

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1150 <WIL>

A:Cross-references: EMBL:Z78015; PIDN:CA801437.1; GSPDB:GN00023; CESP:CS3A5.2

A:Experimental source: clone R02D5

C:Genetics:

A:Gene: CESP:CS3A5.2

A:Map position: 5

A:Introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2; 6

Query Match

Best Local Similarity 64.8%; Score 35; DB 2; Length 1150;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMDYS 11
|:|:|:|:|:|:|
Db 562 VLPVGIDYS 570

Search completed: June 3, 2004, 12:00:01

Job time: 10 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M protein - protein search, using sw model

run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EEVVPXGWDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	70.4	1058	1	CARB_FUSNN
2	36	66.7	102	1	Y1LK_TYDVA
3	36	66.7	1049	1	SPS_ORYSA
4	36	66.7	1068	1	SPS_WAIZE
5	36	66.7	1081	1	SPS2_CRAPL
6	35	64.8	2778	1	FAP_DROME
7	34.5	63.9	748	1	KHL1_HUMAN
8	34	63.0	154	1	PLAS_ORYSA
9	34	63.0	155	1	PLAS_HORVU
10	34	63.0	168	1	SVET_PORNI
11	34	63.0	566	1	SVFB_BORBU
12	34	63.0	908	1	RSL4_ARATH
13	34	63.0	908	1	RPP8_ARATH
14	34	63.0	910	1	RPH8_ARATH
15	33	61.1	276	1	Y939_METJA
16	33	61.1	283	1	PANC_PSEAE
17	33	61.1	394	1	PANC_VIBCH
18	33	61.1	421	1	ACDM_RAT
19	33	61.1	421	1	ECB2_HAUEL
20	33	61.1	423	1	ECB1_HAUEL
21	33	61.1	787	1	ECB2_HUMAN
22	33	61.1	801	1	FGK3_MOUSE
23	33	61.1	806	1	CSK2_CHICK
24	33	61.1	877	1	SULH_SCHPO
25	33	61.1	982	1	ENV_SFVL
26	33	61.1	1401	1	RPOC_VIBCH
27	33	61.1	2717	1	ZEP1_HUMAN
28	32.5	60.2	472	1	ET2A_XENLA
29	32	59.3	97	1	PLAS_DAUCA
30	32	59.3	150	1	YGC1_PSEBK
31	32	59.3	165	1	TPX_LISIN
32	32	59.3	165	1	TRX_LISMO
33	32	59.3	175	1	HES3_RAT

RESULT 1

ID	CARB_FUSNN	STANDARD;	PRT;	1058 AA.
AC	Q8RG86;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
DE	phosphate synthetase ammonia chain).			
GN	CARB OR FN0422.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;			
OC	Fusobacterium.			
OX	NCBI_TaxID=76856;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 25586;			
RX	MEDLINE=2186394; PubMed=1189109;			
RA	Kapattal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein M., Kyrides N., Overbeek R.,			
RA	"Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."			
RL	J. Bacteriol. 184:2005-2018(2002).			
CC	-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.			
CC	-!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	-!- PATHWAY: Arginine biosynthesis.			
CC	-!- PATHWAY: Pyrimidine biosynthesis; first step.			
CC	-!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).			
CC	-!- SIMILARITY: Belongs to the carb family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AE010554; AAL94625.1; ALT_INIT.			
CC	HAMAP; MF_01210; -; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; CPase_L.			
DR	InterPro; IPR005479; CPase_L_D2.			
DR	InterPro; IPR005480; CPase_L_D3.			
DR	InterPro; IPR005481; CPase_L_N.			
DR	InterPro; IPR004362; MGS_like.			
DR	Pfam; PF00289; CFSase_L_chain; 2.			
DR	Pfam; PF02786; CFSase_L_D2; 2.			
DR	Pfam; PF02787; CFSase_L_D3; 1.			

P08203 escherichia
P06190 salmonella
Q9WXR1 thermotoga
Q04827 rattus norv
P30279 homo sapien
P30280 mus musculu
Q90459 brachydanio
P50755 xenopus lae
P49706 gallus gall
P53782 xenopus lae
P55169 gallus gall
P30281 homo sapien

ALIGNMENTS

```

DR Pfam: PF02142; MGS: 1.
DR PRINTS: PR00098; CPSASE.
DR TIGRfam: TIGR01369; CPSaseII_lrg; 1.
DR PROSITE: PS00866; CPSASE_1; 2.
DR PROSITE: PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF7C1E39F CRC64;

Query Match 70.4%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 9.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDYS 11
|:|:|:|:|
DB 190 EIVFNGLYNS 199

RESULT 2
YLIK_TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OC NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92189538; PubMed=1546458;
RA Morris B.A.M.; Richardson K.A.; Haley A.; Zhan X.; Thomas J.E.;
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81103; AAA47947.1; --
CC PIR; A42452; A42452. Geminiv. mov.
CC InterPro; IPR002621; Geminiv. mov.
CC Pfam; PF01708; Geminiv. mov; 1.
CC Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDYS 11
|:|:|:|:|

DR Pfam: PF02142; MGS: 1.
DR PRINTS: PR00098; CPSASE.
DR TIGRfam: TIGR01369; CPSaseII_lrg; 1.
DR PROSITE: PS00866; CPSASE_1; 2.
DR PROSITE: PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF7C1E39F CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMDYS 11
|:|:|:|:|
DB 436 VVFXGMDYS 444

RESULT 4
SPS_MAIZE STANDARD; PRT; 1068 AA.
ID_SPS_MAIZE
AC P31527;
DT 01-JUL-1993 (Rel. 26, Created)

```

01-JUL-1993 (Rel. 26, Last sequence update)
01-JUN-1994 (Rel. 29, Last annotation update)
Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
phosphate glucosyltransferase).
SPS.
Zea mays (Maize).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
PACAD clade; Panicoideae; Andropogoneae; Zea.
NCBI_TaxID=4577;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 71-74; 206-212; 471-481 AND
872-892.
STRAIN=CV. PIONEER 3184; TISSUE=Leaf;
MEDLINE=92338837; PubMed=1840396;
Worrell A.C., Bruneau J.-M., Summerfelt K., Boersig M., Voelker T.A.;
"Expression of a maize sucrose phosphate synthase in tomato alters
leaf carbohydrate partitioning."
Plant Cell 3:1111-1130(1991).
-!- FUNCTION: Involved in the regulation of carbon partitioning in the
leaves of plants. May regulate the synthesis of sucrose and
therefore play a major role as a limiting factor in the export of
photoassimilates out of the leaf.
-!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
sucrose 6-phosphate.
-!- ENZYME REGULATION: Activity regulated by phosphorylation and
moderated by concentration of metabolites and light.
-!- PATHWAY: Sucrose synthesis.
-!- SUBUNIT: Homodimer or homotetramer.
-!- DEVELOPMENTAL STAGE: Germinating seeds or mature leaves.
-!- PTM: Phosphorylated. However, phosphorylation is not essential for
enzyme function.
-!- SIMILARITY: Belongs to the glucosyltransferase family 1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M97550; AAA33513.1; -
PIR; JQ1329; JQ1329.
MaizeDB; 25294; -
InterPro; IPR001296; Glyco_transf_1.
Pfam; PF00534; Glycosyltransferase; Phosphorylation.
DOMAIN 25 31 POLY-GLY.
SEQUENCE 1068 AA; 118575 MW; 07467985E9A1D282 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1068;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 3 VVPXGMDYS 11
|:|:|:|:|
Db 435 VPPGMDFS 443

RESULT 5
SPS2 CRAPL STANDARD; PRT; 1081 AA.
ID SPS2 CRAPL
AC O04933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase 2 (EC 2.4.1.14) (UDP-glucose-fructose-
phosphate glucosyltransferase 2).
GN SPS2.
OS Craterostigma plantagineum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
SPS.

lamiales; Lamiales; Lamiales incertae sedis; Lindernieae;
Craterostigma.
NCBI_TaxID=4153;
[1]
SEQUENCE FROM N.A.
MEDLINE=97451773; PubMed=9306694;
Ingram J., Chandler J.W., Gallagher L., Salamini F., Bartels D.;
"Analysis of cDNA clones encoding sucrose-phosphate synthase in
relation to sugar interconversions associated with dehydration in the
resurrection plant Craterostigma plantagineum Hochst.";
Plant Physiol. 115:113-121(1997).
-!- FUNCTION: Involved in the regulation of carbon partitioning in the
leaves of plants. May regulate the synthesis of sucrose and
therefore play a major role as a limiting factor in the export of
photoassimilates out of the leaf.
-!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
sucrose 6-phosphate.
-!- ENZYME REGULATION: Activity regulated by phosphorylation and
moderated by concentration of metabolites and light.
-!- PATHWAY: Sucrose synthesis.
-!- SUBUNIT: Homodimer or homotetramer (By similarity).
-!- PTM: Phosphorylated. However, phosphorylation is not essential for
enzyme function (By similarity).
-!- SIMILARITY: Belongs to the glucosyltransferase family 1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Y11795; CAA72491.1; -
PIR; T09837; T09837.
InterPro; IPR001296; Glyco_transf_1.
Pfam; PF00534; Glycosyltransferase; Phosphorylation; Multigene family.
Transferase; Glycosyltransferase; Phosphorylation; Multigene family.
DOMAIN 245 248 POLY-SER.
FT DOMAIN 256 264 POLY-GLU.
FT DOMAIN 787 790 POLY-ARG.
SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1081;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 3 VVPXGMDYS 11
|:|:|:|:|
Db 445 VPPGMDFS 453

RESULT 6
FAF_DROME STANDARD; PRT; 2778 AA.
ID FAF_DROME
AC P55824; Q9V9T6; Q9V9T7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
DE (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
GN FAF OR BCDNA:LD22582 OR CG1945.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
RC TISSUE=Eye imaginal disk;
RX MEDLINE=93202020; PubMed=1295747;
Fischer-Vize J.A., Rubin G.M., Lehmann R.;


```

Query Match          63.9%; Score 34.5; DB 1; Length 748;
Best Local Similarity 80.0%; Pred.No.35;
Matches      8; Conservative      1; Mismatches      0; Indels      1; Gaps      1;

QY      1 BEVVPXGMDY 10
      |||||
DB      127 BEVVP-GWDF 135

RESULT 8
PLAS_ORYSA
ID PLAS ORYSA STANDARD; PRT; 154 AA.
AC P20423; Q9SB88;
DT 01-FEB-1991 (Rel. 17, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PETE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OC NCBI_taxID=4530;
RN [1] RN
SEQUENCE FROM N.A.
RP STRAIN=cv. Iipoom; TISSUE=Leaf;
RC Lee J.-S.;
RA "Molecular cloning and characterization of plastocyanin precursor in
RL rice.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN [2] RN
SEQUENCE OF 58-154.
RP STRAIN=cv. Japonica;
RC MEDLINE=9386623; PubMed=2780537;
RA Yano H., Kamo M., Tsugita A., Aso K., Nozu Y.;
RT "The amino acid sequence of plastocyanin from rice (Oryza sativa,
RT subspcies japonica).";
RL Protein Seq. Data Anal. 2:385-389(1989).
CC -!- FUNCTION: Participates in electron transfer between P700 and the
CC cytochrome b6-f complex in photosystem I.
CC -!- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid
CC membrane surface in chloroplasts.
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF093636; AAC78108.1; -.
CC HSSP; P00289; 2PCF.
CC
CC Gramene; P20423; -.
CC InterPro; IPR000923; BlueCu_1.
CC InterPro; IPR01235; Copper blue.
CC InterPro; IPR008972; Cupredoxin.
CC Pfam; PF00127; copper-bind; 1.
CC PRINTS; PR00156; COPPERBLUE.
CC PRODOM; PD001235; Copper blue; 1.
CC PROSITE; PS00196; COPPER_BLUE; 1.
CC Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
CC Transit peptide.
CC
CC TRANSIT 1 57 CHLOROPLAST.
CC CHAIN 58 154 PLASTOCYANIN.
CC DOMAIN 58 154 PLASTOCYANIN-LIKE.
CC METAL 94 94 COPPER (BY SIMILARITY).
CC METAL 139 139 COPPER (BY SIMILARITY).
CC METAL 142 142 COPPER (BY SIMILARITY).
CC METAL 147 147 COPPER (BY SIMILARITY).
CC SEQUENCE 154 AA; 15577 MW; E45725D25B5F400D CRC64;
CC

```

Query Match 63.0%; Score 34; DB 1; Length 154;
 Best Local Similarity 54.5%; Pred. No. 8.7;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
 : : : : :
 DB 100 EDAVPSGVDVS 110

RESULT 9

PLAS_HORVU PLAS_HORVU STANDARD; PRT; 155 AA.
 AC P08248;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plastocyanin, chloroplast precursor.
 GN PTE.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Bomi;
 RA Nielsen O.S.; Gausing K.;
 RT "The precursor of barley plastocyanin: sequence of cDNA clones and
 gene expression in different tissues";
 RL FEBS Lett. 225:159-162(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. NK 1558;
 RX MEDLINE=94039081; PubMed=8223592;
 RA Nielsen P.; Gausing K.;
 RT "In vitro binding of nuclear proteins to the barley plastocyanin gene
 promoter region";
 RL Eur. J. Biochem. 217:97-104(1993).
 CC -1- FUNCTION: Participates in electron transfer between P700 and the
 cytochrome b6-f complex in photosystem I.
 CC -1- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid
 membrane surface in chloroplasts.
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00704; CAA69696.1; -;
 DR EMBL; Z28347; CAA82201.1; -;
 DR PIR; S38255; S38255.
 DR HSP; P00289; 2PCF.
 DR InterPro; IPR000923; BlueCu 1.
 DR InterPro; IPR001235; Copper blue.
 DR InterPro; IPR008972; Cupredoxin.
 DR Pfam; PF00127; copper-bind; 1.
 DR PRINTS; PR00156; COPPERBLUE.
 DR ProDom; PD001235; Copper blue; 1.
 DR PROSITE; PS00196; COPPER_BLUE; 1.
 DR Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
 KW Transit peptide.
 FT TRANSIT 1 58 CHLOROPLAST.
 FT CHAIN 59 155 PLASTOCYANIN
 FT DOMAIN 59 155 PLASTOCYANIN-LIKE
 FT METAL 95 95 COPPER (BY SIMILARITY).
 FT METAL 140 140 COPPER (BY SIMILARITY).
 FT METAL 143 143 COPPER (BY SIMILARITY).
 FT METAL 148 148 COPPER (BY SIMILARITY).

FT VARIANT 120 120 T -> N (IN CV. NK 1558).
 SQ SEQUENCE 155 AA; 15709 MW; DAA7EAB5F6F4F91 CRC64;
 Query Match 63.0%; Score 34; DB 1; Length 155;
 Best Local Similarity 54.5%; Pred. No. 8.7;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
 : : : : :
 DB 101 EDAVPSGVDVS 111

RESULT 10

PLAT_POPNI PLAT_POPNI STANDARD; PRT; 168 AA.
 AC P11970;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plastocyanin B, chloroplast precursor.
 GN PTE.
 OS Populus nigra (Lombardy poplar).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid J; Malpighiales; Salicaceae; Salicaceae; Populus.
 OC NCBI_TaxID=3691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Italica; TISSUE=Leaf;
 RA Reichert J.; Jenzelewski V.; Haehnel W.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 70-168.
 RC STRAIN=cv. Italica;
 RA Dimitrov M.I.; Egorov C.A.; Donchev A.A.; Atanasov B.P.;
 RL "Complete amino acid sequence of poplar plastocyanin b.";
 FEBS Lett. 226:17-22(1987).
 CC -1- FUNCTION: Participates in electron transfer between P700 and the
 cytochrome b6-f complex in photosystem I.
 CC -1- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid
 membrane surface in chloroplasts.
 CC -1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF
 POPLAR PLASTOCYANINS A AND B.
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z50186; CAA90565.1; -;
 DR PIR; S00210; S00210.
 DR PIR; S58208; S58208.
 DR HSP; P00299; IPLC.
 DR InterPro; IPR000923; BlueCu 1.
 DR InterPro; IPR001235; Copper blue.
 DR InterPro; IPR008972; Cupredoxin.
 DR Pfam; PF00127; copper-bind; 1.
 DR PRINTS; PR00156; COPPERBLUE.
 DR ProDom; PD001235; Copper blue; 1.
 DR PROSITE; PS00196; COPPER_BLUE; 1.
 DR Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
 KW Transit peptide; Multigene family.
 FT TRANSIT 1 69 CHLOROPLAST.
 FT CHAIN 70 168 PLASTOCYANIN B.
 FT DOMAIN 70 168 PLASTOCYANIN-LIKE.
 FT METAL 106 106 COPPER.
 FT METAL 133 153 COPPER.
 FT METAL 156 156 COPPER.
 FT METAL 161 161 COPPER.


```
3Q SEQUENCE 168 AA; 16981 MW; F20DA6EA2038A8EA CRC64;
Query Match 63.0%; Score 34; DB 1; Length 168;
Best Local Similarity 54.5%; Pred. No. 9.5;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

2Y 1 BEVVPXGMDYS 11
|:|:|:|
3b 112 EDVPSGVDVS 122

RESULT 11
SYFB_BORBU STANDARD; PRT; 566 AA.
AC P94283;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PheRS).
EN PHE1 OR B50514.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=119;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Barbour A.G., Hinnebusch J.;
RA "phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and
RT thiorodoxin reductase gene of Borrelia burgdorferi.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586 (1997).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
CC family. Subfamily 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```
SQ SEQUENCE 566 AA; 65173 MW; 9D48C8B5D6D3B74B CRC64;
Query Match 63.0%; Score 34; DB 1; Length 566;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPXGMDY 10
|:|:|:|
Db 169 VPFGMDY 175

RESULT 12
REL4_ARATH STANDARD; PRT; 908 AA.
AC Q9FJK8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and PAC clones.";
RL DNA Res. 5:297-308 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayaishizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Potential disease resistance protein.
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (By similarity).
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME=NB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW=http://niblrts.ucdavis.edu".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


```



```

FT DOMAIN 146 459 NB-ARC.
FT REPEAT 575 599 LRR 1.
FT REPEAT 600 623 LRR 2.
FT REPEAT 842 867 LRR 3.
FT NP BIND 192 199 ATP (POTENTIAL).
SQ SEQUENCE 908 AA; 104448 MW; 3111991B17239693 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 908;
Best Local Similarity 60.0%; Pred. No. 54;
Matches 5; Conservative 2; Mismatches 0; Gaps 0;

Qy 1 EEVVPXGMDY 10
   |::|||
Db 883 EKLVPQGEDY 892

RESULT 13
RPP8_ARATH STANDARD; PRT; 908 AA.
AC Q8W4J9; Q8GWG5; Q8W5A1; Q9ZSY3; Q9ZSY4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Disease resistance protein RPP8 (Resistance to Peronospora parasitica protein 8).
GN RPP8 OR HRT OR AT5G43470 OR MWF20.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND RPP8-3, AND VARIANTS.
RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
RX MEDLINE=99030193; PubMed=9811794;
RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis.";
RL Plant Cell 10:1861-1874 (1998).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
RC STRAIN=cv. Di-17;
RX MEDLINE=2021766; PubMed=10810142;
RA Cooley M.B., Pathirana S., Wu H.-J., Kachroo P., Kleissig D.F.;
RT "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens.";
RL Plant Cell 12:663-676 (2000).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones.";
RL DNA Res. 7:31-63 (2000).
RN [4]
SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yanada K., Lim J., Dale J.M., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Southwick A.M., Wu H.C., Kim C.J., Sakano H., Wu T., Yu G.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Lee J.M., Toriumi M.J.,
RA Miranda M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,

```

```

RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vayenberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis genome.";
RL Science 302:842-846 (2003).
RN [5]
SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.
RN [6]
INTERACTION WITH TIP.
RX MEDLINE=20496823; PubMed=11041886;
RA Ren T., Qu F., Morris T.J.;
RT "HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus.";
RL Plant Cell 12:1917-1926 (2000).
CC -!- FUNCTION: Disease resistance protein. Resistance proteins guard the plant against pathogens that contain an appropriate avirulence protein via an indirect interaction with this avirulence protein. That triggers a defense system including the hypersensitive response, which restricts the pathogen growth. The interaction with TIP (TCV-interacting protein) may be essential for the recognition of the avirulence proteins, and the triggering of the defense response.
CC -!- SUBUNIT: Interacts with the NAC protein TIP.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8W4J9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8W4J9-2; Sequence=VSP_007171, VSP_007172;
CC Note=Has been shown to exist only in cv. Columbia so far;
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of pathogen recognition.
CC -!- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and cv. Columbia are probably due to an unequal crossing-over between the highly related RPP8 and RPP8A genes present in cv. Landsberg erecta. Such variations probably modify the specificity of pathogen recognition.
CC -!- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8 specifically recognizes the Emco5 avirulence protein from Peronospora parasitica, while it is not the case in cv. Di-17, where it confers resistance to Turnip Crinkle Virus upon recognition of the viral capsid protein.
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family. RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene homologs;
WWW="http://niblr.ucdavis.edu".
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
-----
EMBL; AF089710; AAC83165.1; -
EMBL; AF089711; AAC78631.1; -
EMBL; AF234174; AAF36987.1; -
EMBL; AB025636; BAA97426.1; -
EMBL; AY062514; AAL32592.1; -
EMBL; AK118862; BAC43449.1; -
InterPro; IPR000767; Disease_resist.

```

```

DR InterPro; IPR001611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 2.
DR Pfam; PF00931; NB-ARC; 1.
DR PRINTS; PR00364; DISEASERSIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat;
KW Alternative splicing; Polymorphism.
FT DOMAIN 10 45
FT LEUCINE-ZIPPER.
FT NP_BIND 146 459
FT REPEAT 600 623
FT REPEAT 842 867
FT NP_BIND 192 199
FT VARSPLIC 294 308
FT
FT
FT VARSPLIC 309 908
FT
FT VARIANT 27 31
FT VARIANT 29 29
FT VARIANT 85 87
FT
FT VARIANT 88 91
FT VARIANT 100 100
FT VARIANT 121 121
FT VARIANT 129 129
FT VARIANT 133 133
FT VARIANT 138 138
FT VARIANT 173 173
FT VARIANT 177 177
FT VARIANT 253 253
FT VARIANT 257 257
FT VARIANT 265 265
FT VARIANT 267 267
FT VARIANT 270 270
FT VARIANT 283 283
FT VARIANT 369 369
FT
FT VARIANT 387 387
FT VARIANT 399 399
FT
FT VARIANT 426 426
FT
FT VARIANT 429 429
FT
FT VARIANT 436 446
FT
FT VARIANT 438 443
FT VARIANT 450 450
FT VARIANT 460 460
FT VARIANT 480 481
FT VARIANT 485 486
FT VARIANT 489 489
FT VARIANT 514 514
FT
FT VARIANT 518 518
FT
FT VARIANT 519 519
FT VARIANT 528 528
FT VARIANT 537 537
FT VARIANT 542 542
FT VARIANT 543 543
FT VARIANT 547 550
FT VARIANT 549 549
FT VARIANT 558 565
FT
FT VARIANT 558 560
FT VARIANT 564 564
FT VARIANT 584 584
FT VARIANT 595 595
FT

```

```

Query Match 63.0%; Score 34; DB 1; Length 908;
Best Local Similarity 60.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMDY 10
DB 883 EKLVPGGEDY 892
|::|||
|::|||

RESULT 14
ID RPH8_ARATH STANDARD; PRT; 910 AA.
AC P59584;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Disease resistance protein RPH8A (RPP8 homolog A).
GN RPH8A.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=99030193; PubMed=9811794;
RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intragenic recombination and diversifying selection contribute to the
RT evolution of downy mildew resistance at the RPP8 locus of
RT Arabidopsis";
RL Plant Cell 10:1861-1874 (1998).
CC -!- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. In contrast to
CC RPP8, it does not specifically recognize the Emco5 avirulence
CC protein from Peronospora parasitica.
CC -!- MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
CC present due to an unequal crossing over between the RPP8 and RPH8A
CC genes that creates a unique RPP8 gene.
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME-NIB-LRRS;
CC NOTS=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrns.ucdavis.edu".
CC InterPro; IPR000767; Disease_resist.
CC InterPro; IPR001611; LRR.
CC InterPro; IPR002182; NB-ARC.
CC Pfam; PF00560; LRR; 2.
CC Pfam; PF00931; NB-ARC; 1.
CC PRINTS; PR00364; DISEASERSIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT DOMAIN 10 45
FT LEUCINE-ZIPPER.
FT NP_BIND 146 459
FT REPEAT 602 625
FT REPEAT 844 863
FT NP_BIND 192 199
FT ATP (POTENTIAL).
SQ SEQUENCE 910 AA; 105263 MW; 5B1E9F65A19A12EE CRC64;

Query Match 63.0%; Score 34; DB 1; Length 910;
Best Local Similarity 60.0%; Pred. No. 55;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGMDY 10
DB 885 EKLVPGGEDY 894
|::|||
|::|||

```

```
RESULT 15
Y939_METJA STANDARD; PRT; 276 AA.
AC Q58349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0939.
GN MJ0939
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
EX MEDLINE=96337999; PubMed=8688087;
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Zhou L., Fleischmann R.D.,
RA Sullivan G.G., White O., Olsen G.J., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Corton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kierk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67537; BAB98946.1; -.
DR PIR; C64417; C64417.
DR TIGR; MJ0939; -.
DR InterPro; IPR001163; SRNP Sm.
DR InterPro; IPR005358; UPF0153.
DR Pfam; PF03692; UPF0153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 33454 MW; 97BD69D392BC8FDF CRC64;

Query Match 61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred.No. 26;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
Db 141 EELIENGMEHS 151
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 BEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP TREMBL 25:*

2: sp archaea:*

3: sp bacteria:*

4: sp fungi:*

5: sp human:*

6: sp invertebrate:*

7: sp mammal:*

8: sp mhc:*

9: sp organelle:*

10: sp phage:*

11: sp plant:*

12: sp rodent:*

13: sp virus:*

14: sp vertebrate:*

15: sp unclassified:*

16: sp viirus:*

17: sp bacteriap:*

18: sp archaeap:*

19: sp archaeap:*

20: sp archaeap:*

21: sp archaeap:*

22: sp archaeap:*

23: sp archaeap:*

24: sp archaeap:*

25: sp archaeap:*

26: sp archaeap:*

27: sp archaeap:*

28: sp archaeap:*

29: sp archaeap:*

30: sp archaeap:*

31: sp archaeap:*

32: sp archaeap:*

33: sp archaeap:*

34: sp archaeap:*

35: sp archaeap:*

36: sp archaeap:*

37: sp archaeap:*

38: sp archaeap:*

39: sp archaeap:*

40: sp archaeap:*

41: sp archaeap:*

42: sp archaeap:*

43: sp archaeap:*

44: sp archaeap:*

45: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	Q12479	saccharomyc
2	38	70.4	319	Q8ESV7	Q8ESV7 oceanobacil
3	38	70.4	363	Q30260	Q30260 archaeoglob
4	38	70.4	1044	Q8DIH0	Q8DIH0 synchococ
5	36	66.7	149	Q82B34	Q82B34 enterococc
6	36	66.7	341	Q22081	Q22081 citrus unsh
7	36	66.7	348	Q22096	Q22096 citrus unsh
8	36	66.7	452	Q8W568	Q8W568 arabidopsis
9	36	66.7	460	Q9C9T7	Q9C9T7 arabidopsis
10	36	66.7	1047	P93782	P93782 saccharum o
11	36	66.7	1083	Q9SN30	Q9SN30 arabidopsis
12	36	66.7	1084	Q43010	Q43010 oryza sativ
13	36	66.7	1100	Q8S064	Q8S064 oryza sativ
14	35	64.8	219	Q9GQ04	Q9GQ04 eriocheir s
15	35	64.8	253	Q8XPAS	Q8XPAS clostridium
16	35	64.8	298	O52367	O52367 thizobium t

17	35	64.8	308	16	Q8SCC0	Q8SCC0 pseudomonas
18	35	64.8	368	2	Q8KL43	Q8KL43 rhizobium e
19	35	64.8	425	5	Q9XVK4	Q9XVK4 caenorhabdi
20	35	64.8	433	16	Q9A382	Q9A382 caulobacter
21	35	64.8	440	17	Q9YFI3	Q9YFI3 aeropyrum p
22	35	64.8	474	16	Q83L92	Q83L92 shigella fl
23	35	64.8	511	2	O52880	O52880 escherichia
24	35	64.8	517	16	Q8XZL5	Q8XZL5 raistonia s
25	35	64.8	595	10	Q946J7	Q946J7 androgaphi
26	35	64.8	745	5	Q9SP46	Q9SP46 carinus na
27	35	64.8	1031	5	Q9UEA3	Q9UEA3 callinectes
28	35	64.8	1150	5	O17704	O17704 caenorhabdi
29	35	64.8	1410	2	O52673	O52673 escherichia
30	35	64.8	1420	2	O52666	O52666 escherichia
31	35	64.8	1474	17	O27146	O27146 methanobact
32	35	64.8	1828	16	Q98K29	Q98K29 rhizobium l
33	34.5	63.9	748	4	Q8TEJ7	Q8TEJ7 homo sapien
34	34	63.0	215	16	Q8R9L5	Q8R9L5 thermaanaer
35	34	63.0	222	16	Q8E5L9	Q8E5L9 streptococc
36	34	63.0	222	16	Q8DZW9	Q8DZW9 streptococc
37	34	63.0	245	16	Q7V6Q4	Q7V6Q4 prochloroco
38	34	63.0	251	9	Q85SK6	Q85SK6 mycobacteri
39	34	63.0	284	2	Q8GP33	Q8GP33 lactobacilli
40	34	63.0	290	16	Q8U7J0	Q8U7J0 agrobacteri
41	34	63.0	296	17	Q9YET8	Q9YET8 aeropyrum p
42	34	63.0	337	16	Q7UF95	Q7UF95 rhodospirell
43	34	63.0	344	16	Q815A7	Q815A7 bacillus ce
44	34	63.0	357	17	O29920	O29920 archaeoglob
45	34	63.0	366	17	O29451	O29451 archaeoglob

ALIGNMENTS

RESULT 1

Q12479	PRELIMINARY;	PRT;	156 AA.
ID Q12479;			
AC Q12479;			
DT 01-NOV-1996 (TREMBLrel. 01, Created)			
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE ORF YOR013W.			
GN YOR013W.			
OS Saccharomyces cerevisiae (Baker's yeast).			
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae; Saccharomycetes.			
OX NCBI_TaxID=4932;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA De haan M., Grivell L.A., Maarse A.C.;			
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA MIPS;			
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN=FY1679;			
RA De haan M., Maarse A.C., Grivell L.A.;			
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
RN [4]			
RP SEQUENCE FROM N.A.			
RC STRAIN=FY1679;			
RA MEDLINE=9401918; PubMed=9413243;			
RL Dumont M.E., Schlichter J.B., Cardillo T.S., Hayes M.K., Bethlenny G.,			
RT Sherman F.;			
RT "CYC2 encodes a factor involved in mitochondrial import of yeast			
RL Mol. Cell. Biol. 13:6442-6451 (1993).			
RN [5]			
RP SEQUENCE FROM N.A.			
RC STRAIN=FY1679;			
RA MEDLINE=94169519; PubMed=7764548;			

```

RA Lee Y.S., Shimizu J., Yoda K., Yamasaki M.;
RT "Molecular cloning of a gene, DHS1, which complements a drug-
RT hypersensitive mutation of the yeast Saccharomyces cerevisiae.";
RL Biosci. Biotechnol. Biochem. 58:391-395(1994).
DR EMBL; Z74920; CAA39201.1; -.
DR EMBL; X87331; CAA60762.1; -.
DR PIR; S54619; S54619.
DR SGD; S0005539; YOR013W.
SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;

Query Match 74.1%; Score 40; DB 3; Length 156;
Best Local Similarity 77.8%; Pred. No. 3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMDY 10
DB 50 EVNPLGMDY 58
|||||

RESULT 2
Q8ESV7 PRELIMINARY; PRT; 319 AA.
AC Q8ESV7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical conserved protein.
GN O80509.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004594; BAC12465.1; -.
DR InterPro; IPR001279; Blackmase-like.
DR Pfam; PF00753; lactamase_B; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDA54BF13E79E37 CRC64;

Query Match 70.4%; Score 38; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVPXGMDY 10
DB 189 EQLVPHGIDY 198
|||||

RESULT 3
O30260 PRELIMINARY; PRT; 363 AA.
AC O30260;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Af2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=22334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

```

```

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.W.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001109; AAB91255.1; -.
DR PIR; D69551; D69551.
DR TIGR; AF2411; -.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;

Query Match 70.4%; Score 38; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
DB 120 ENIVFYGIDFS 130
|||||

RESULT 4
Q8DIH0 PRELIMINARY; PRT; 1044 AA.
AC Q8DIH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN T11618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpou S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BAC09170.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001036; Acflavin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRfams; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 70.4%; Score 38; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
DB 843 BEVLPNGIGYS 853
|||||

```

```

RESULT 5
Q822B4
ID Q822B4 PRELIMINARY; PRT; 149 AA.
AC Q822B4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Large conductance mechanosensitive channel protein.
EN MSCL OR EF3152.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Iran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Frazer C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016956; AA082828.1; -.
DR TIGR; EF3152; -.
DR GO; GO:0016020; C:membrane, IEA.
DR GO; GO:005216; P:ion channel activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001185; MS_channel.
DR Pfam; PF01741; MsCL; 1.
DR PRINTS; PR01264; MECHCHANNEL.
DR ProDom; PD007253; MS_channel; 1.
DR TIGRFAMs; TIGR00220; mscl; 1.
KW Complete proteome.
SQ SEQUENCE 149 AA; 16127 MW; 555799BF1E47D34E CRC64;

Query Match 66.7%; Score 36; DB 16; Length 149;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGMDY 10
DB 113 EVVFTSEDY 122

RESULT 6
Q22081
ID Q22081 PRELIMINARY; PRT; 341 AA.
AC Q22081;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
EN CitSPS2.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurooids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; AB006660; BAA22071.1; -.
DR PIR; S72650; S72650.
FT NON TER 1
FT NON TER 348
SQ SEQUENCE 348 AA; 38556 MW; BE1C21EBA6FF5C5E CRC64;

Query Match 66.7%; Score 36; DB 10; Length 348;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 234 VIPPGMDFS 242

RESULT 8
Q8W568
ID Q8W568 PRELIMINARY; PRT; 452 AA.
AC Q8W568;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Atig73750/F25P22_17.
EN CitSPS2.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurooids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Gen. Genet. 252:346-351(1996).
RN [2]
SEQUENCE FROM N.A.

```

```

RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; AB006319; BAA23215.1; -.
DR PIR; S72649; S72649.
FT NON TER 1
FT NON TER 341
SQ SEQUENCE 341 AA; 38136 MW; 61417A69C4560777 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 341;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 228 VIPPGMDFS 236

RESULT 7
Q22096
ID Q22096 PRELIMINARY; PRT; 348 AA.
AC Q22096;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
EN CitSPS3.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurooids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Gen. Genet. 252:346-351(1996).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178(1999).
DR EMBL; AB006660; BAA22071.1; -.
DR PIR; S72650; S72650.
FT NON TER 1
FT NON TER 348
SQ SEQUENCE 348 AA; 38556 MW; BE1C21EBA6FF5C5E CRC64;

Query Match 66.7%; Score 36; DB 10; Length 348;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 234 VIPPGMDFS 242

RESULT 8
Q8W568
ID Q8W568 PRELIMINARY; PRT; 452 AA.
AC Q8W568;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Atig73750/F25P22_17.
EN CitSPS2.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurooids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Gen. Genet. 252:346-351(1996).
RN [2]
SEQUENCE FROM N.A.

```

OS Arabidopsis thaliana (mouse-ear cress);
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RA "Arabidopsis cDNA clones";
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Shinn M., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Iehida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RA "Arabidopsis ORF clones";
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL
 DR EMBL; AF419606; AAL31938.1; -;
 DR EMBL; AY113044; AAM47352.1; -;
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0004287; F:prolyl oligopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estrs.
 DE PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 SQ SEQUENCE 452 AA; 45682 MW; A159955B21742C4A CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 452;
 Best Local Similarity 70.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDY 10
 Db 210 EEDVPSANDY 219
 ID Q9C9T7 PRELIMINARY; PRT; 460 AA.
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN F25p22.17.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen E., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu X.X., Luros J.S., Maiti R., Marziali A.,
 RA Millscher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RT Nature 408:816-820 (2000).
 RL EMBL; AC013679; AGS2073.1; -;
 DR PIR; G96764; G96764.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0004287; F:prolyl oligopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estrs.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 460 AA; 50564 MW; E94B27B5C4B249EC CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 460;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDY 10
 Db 218 EEDVPSANDY 227
 ID P93782 PRELIMINARY; PRT; 1047 AA.
 AC P93782;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (Fragment).
 GN SOSP81.
 OS Saccharum officinarum (Sugarcane).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACAD clade; Panicoideae; Andropogoneae; Saccharum.
 CX NCBI_TaxID=4547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Sugiharto B., Sakakibara H., Sugiyama T.;
 RT "Differential Expression of Two Genes for Sucrose-Phosphate Synthase
 in Sugarcane: Molecular Cloning of the cDNAs and Comparative Analysis
 of Gene Expression";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB001337; BAA19241.1; -;
 DR GO; GO:004524; F:sucrose-phosphate synthase activity; IEA.
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycos transf 1; 1.
 DR Glycoyltransferase; Transferase.
 FT NON_TER 1
 SQ SEQUENCE 1047 AA; 116379 MW; D0EDB34961E1D83D CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 1047;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11
 Db 414 VVPXGMDYS 422
 RESULT 11

Q9SN30
ID Q9SN30 PRELIMINARY; PRT; 1093 AA.
AC Q9SN30;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
GN F28M11.40 OR AT4G10120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lencke K., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049487; CAB39764.1; -
DR EMBL; AL161516; CAB78135.1; -
DR PIR; T04062; T04062.
DR GO; GO:0045524; P:sucrose-phosphate synthase activity; IEA.
DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001296; Glyco trans 1.
DR Pfam; PF00534; Glycos transf 1; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1093 AA; 12268 MW; EFCL26FCA2137BB1 CRC64;
Query Match 66.7%; Score 36; DB 10; Length 1093;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
DB 493 VIPPGMDFS 491
RESULT 12
Q43010 PRELIMINARY; PRT; 1084 AA.
ID Q43010;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose phosphate synthase.
GN SPS1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. japonica;
RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shimada H.,
RA Fujimura T.;
RT "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS)
RT gene that is specifically expressed in the source organ."
RL Plant Sci. 112:207-217(1995).
DR EMBL; D45890; BAA08304.1; -
DR PIR; T04103; T04103.
DR Gramene; Q43010; -
DR GO; GO:0004459; P:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.

DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycos transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1084 AA; 119417 MW; 4D0D79AFEDD2F92B CRC64;
Query Match 66.7%; Score 36; DB 10; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
DB 453 VIPPGMDFS 461
RESULT 13
Q8S064 PRELIMINARY; PRT; 1100 AA.
ID Q8S064;
AC Q8S064;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative sucrose-phosphate synthase.
GN P0678F11.14.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0678F11.1";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003437; BAE86107.1; -
DR Gramene; Q8S064; -
DR GO; GO:0004459; P:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001296; Glyco transf_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycos transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1100 AA; 121170 MW; 084F4604BA389CAD CRC64;
Query Match 66.7%; Score 36; DB 10; Length 1100;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
DB 469 VIPPGMDFS 477
RESULT 14
Q9GQ04 PRELIMINARY; PRT; 219 AA.
ID Q9GQ04;
AC Q9GQ04;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Na+/K+/2Cl-cotransporter (Fragment).
OS Eriocheir sinensis (Chinese mitten crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Grapsoidea; Varunidae; Eriocheir.
OX NCBI_TaxID=95602;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gill;

RA Wehrauch D., Towle D.W.;
RT "Na+/H+-exchanger and Na+/K+/2Cl- -cotransporter are expressed in
RT gills of the euryhaline Chinese crab *Eriocheir sinensis*,"
EL Comp. Biochem. Physiol. 126:S158-S158(2000).
DR EMBL: AF301160; AAC39938.1; -.
FI NON_TER 219 219
SQ SEQUENCE 219 AA; 24159 MW; 599442DA26F0D3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPKGMQYS 11
DB 107 VPQGLDYS 114

RESULT 15

Q8XPA8 PRELIMINARY; PRT; 253 AA.
AC Q8XPA8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein CPE0057.
GN CPE0057.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003185; BAB79763.1; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR009036; MoEB.
DR InterPro: IPR000205; NAD_BS.
DR InterPro: IPR000594; ThiF_domain.
DR Pfam: PF00899; ThiF; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 27714 MW; 814DF79D3E0D7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
DB 108 EEIIPDDVDY 117

Search completed: June 3, 2004, 11:57:33
Job time : 29.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 BEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	11	5	ABB80561 Hepatitis
2	52	96.3	11	5	ABB80524 Hepatitis
3	52	96.3	11	5	ABB80529 Hepatitis
4	52	96.3	11	5	ABB80528 Hepatitis
5	52	96.3	11	5	ABB80562 Hepatitis
6	47	87.0	11	5	ABB80542 Hepatitis
7	47	87.0	11	5	ABB80543 Hepatitis
8	47	87.0	11	5	ABB80538 Hepatitis
9	46	85.2	11	5	ABB80525 Hepatitis
10	46	85.2	11	5	ABB80548 Hepatitis
11	46	85.2	11	5	ABB80521 Hepatitis
12	46	85.2	11	5	ABB80522 Hepatitis
13	46	85.2	11	5	ABB80547 Hepatitis
14	46	85.2	11	5	ABB80566 Hepatitis
15	46	85.2	11	5	ABB80556 Hepatitis
16	46	85.2	11	5	ABB80557 Hepatitis
17	46	85.2	11	5	ABB80551 Hepatitis
18	46	85.2	11	5	ABB80563 Hepatitis
19	46	85.2	11	5	ABB80565 Hepatitis
20	46	85.2	11	5	ABB80567 Hepatitis
21	46	85.2	11	5	ABB80559 Hepatitis
22	46	85.2	11	5	ABB80526 Hepatitis
23	46	85.2	11	5	ABB80564 Hepatitis
24	46	85.2	11	5	ABB80568 Hepatitis
25	45	83.3	11	5	ABB80523 Hepatitis

26	45	83.3	11	5	ABB80558	Abb80558 Hepatitis
27	45	83.3	11	5	ABB80560	Abb80560 Hepatitis
28	45	83.3	11	5	ABB80527	Abb80527 Hepatitis
29	44	81.5	11	5	ABB80534	Abb80534 Hepatitis
30	44	81.5	11	5	ABB80533	Abb80533 Hepatitis
31	41	75.9	11	5	ABB80536	Abb80536 Hepatitis
32	41	75.9	11	5	ABB80535	Abb80535 Hepatitis
33	41	75.9	11	5	ABB80540	Abb80540 Hepatitis
34	41	75.9	11	5	ABB80539	Abb80539 Hepatitis
35	40	74.1	11	5	ABB80549	Abb80549 Hepatitis
36	40	74.1	11	5	ABB80544	Abb80544 Hepatitis
37	40	74.1	11	5	ABB80537	Abb80537 Hepatitis
38	40	74.1	11	5	ABB80541	Abb80541 Hepatitis
39	40	74.1	11	5	ABB80553	Abb80553 Hepatitis
40	40	74.1	11	5	ABB80552	Abb80552 Hepatitis
41	40	74.1	11	5	ABB80545	Abb80545 Hepatitis
42	39	72.2	11	5	ABB80546	Abb80546 Hepatitis
43	39	72.2	11	5	ABB80554	Abb80554 Hepatitis
44	39	72.2	11	5	ABB80550	Abb80550 Hepatitis
45	39	72.2	11	5	ABB80555	Abb80555 Hepatitis

ALIGNMENTS

RESULT 1
ID ABB80561 standard; peptide; 11 AA.
XX ABB80561;
XX AC ABB80561;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX Synthetic.
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX
SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
| | | | | | | | | |
Db 1 EEVVPXGMDYS 11

RESULT 2
ABB80524
ID ABB80524 standard; peptide; 11 AA.

AC ABB80524;
XX
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.

PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX

SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
| | | | | | | | | |
Db 1 EEVVPXGMDYS 11

RESULT 3
ABB80529
ID ABB80529 standard; peptide; 11 AA.

AC ABB80529;
XX
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX

SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
 |||||
 DB 1 EEVVPXGMDYS 11

RESULT 4

ABB80528

ID ABB80528 standard; peptide; 11 AA.
 XX
 AC ABB80528;
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 PD 31-JAN-2002.
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.

XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 PS Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

XX
 PS Claim 17; Page 64; 69pp; English.
 CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

XX
 SQ Sequence 11 AA;
 Query Match 96.3%; Score 52; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
 |||||
 DB 1 EEVVPXGMDYS 11

RESULT 5

ABB80562

ID ABB80562 standard; peptide; 11 AA.

XX
 AC ABB80562;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

XX OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"

XX
 PN WO200208251-A2.
 PD 31-JAN-2002.

XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.

XX
 PA (CORV-) CORVAS INT INC.

XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.

XX
 PS Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
 XX
 PS Claim 17; Page 65; 69pp; English.

XX
 CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

XX
 SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
 |||||
 DB 1 EEVVPXGMDYS 11

RESULT 6

ABB80542

ID ABB80542 standard; peptide; 11 AA.

XX
 AC ABB80542;

XX
 DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 KW Synthetic.
 OS
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 PN WO200208251-A2.
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 XX PF
 XX 21-JUL-2000; 2000US-0220101P.
 XX PR
 XX (CORV-) CORVAS INT INC.
 XX PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX PI
 XX WPI; 2002-361643/39.
 XX DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C virus
 XX protease.
 XX PT
 XX Claim 17; Page 65; 69pp; English.
 XX PS
 XX The sequence represents a peptide compound of the invention having
 XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 XX invention are alpha-ketoamide peptide analogues. The peptides have
 XX virucide activity, and are useful for treating and in the manufacture of
 XX a medicament to treat disorders associated with HCV protease. A
 XX pharmaceutical composition comprising the peptide as an active ingredient
 XX is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB 1 EEVVPXGQDYS 11
 RESULT 7
 ABB80543
 ID ABB80543 standard; peptide; 11 AA.
 AC ABB80543;
 XX
 XX 08-OCT-2002 (first entry)
 DT
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 KW Synthetic.
 OS
 XX

FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 PN WO200208251-A2.
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 XX PF
 XX 21-JUL-2000; 2000US-0220101P.
 XX PR
 XX (CORV-) CORVAS INT INC.
 XX PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX PI
 XX WPI; 2002-361643/39.
 XX DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C virus
 XX protease.
 XX PT
 XX Claim 17; Page 65; 69pp; English.
 XX PS
 XX The sequence represents a peptide compound of the invention having
 XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 XX invention are alpha-ketoamide peptide analogues. The peptides have
 XX virucide activity, and are useful for treating and in the manufacture of
 XX a medicament to treat disorders associated with HCV protease. A
 XX pharmaceutical composition comprising the peptide as an active ingredient
 XX is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 87.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB 1 EEVVPXGQDYS 11
 RESULT 8
 ABB80538
 ID ABB80538 standard; peptide; 11 AA.
 AC ABB80538;
 XX
 XX 08-OCT-2002 (first entry)
 DT
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 KW Synthetic.
 OS
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT

FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX WO200208251-A2.
PN
PD
PD 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
XX
XX Query Match 87.0%; Score 47; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.023;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB |||||
1 EEVVPXGQDYS 11
|||
RESULT 9
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
XX ABB80525;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
PN
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
XX
XX Query Match 85.2%; Score 46; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.036;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB |||||
1 EEVVPXGMSYS 11
|||
RESULT 10
ABB80548
ID ABB80548 standard; peptide; 11 AA.
XX
XX ABB80548;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
PN
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX 31-JAN-2002.

XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
SQ

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11

RESULT 11
ABB80521
ID ABB80521 standard; peptide; 11 AA.
XX
XX ABB80521;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

Query Match 85.2%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11

RESULT 12
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
XX ABB80522;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

Query Match 85.2%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.036;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

YV 1 EEVVPXGMDYS 11
 |||||
 bB 1 EEVVPXGMSYS 11
 |||||

RESULT 13
 ABB80547
 ID ABB80547 standard; peptide; 11 AA.
 AC ABB80547;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Query Match 85.2%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.036;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

YV 1 EEVVPXGMDYS 11
 |||||
 bB 1 EEVVPXGTDYS 11
 |||||

RESULT 14
 ABB80566
 ID ABB80566 standard; peptide; 11 AA.
 AC ABB80566;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "2-aminoisobutyl carbonyl residue forming a keto
 amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US023169.
 XX 21-JUL-2000; 2000US-0220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 FT activity useful for treating disorders associated with hepatitis C virus
 FT protease.
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Query Match 85.2%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.036;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

YV 1 EEVVPXGMDYS 11
 |||||
 bB 1 EEVVPXGMSYS 11
 |||||

RESULT 15
 ABB80556
 ID ABB80556 standard; peptide; 11 AA.
 AC ABB80556;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #36.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 OS Synthetic.
 XX
 XX
 FH Location/Qualifiers
 FT Key 1
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT Modified-site 11 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 XX
 XX WO200208251-A2.
 PN
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX
 XX WPI; 2002-361643/39.
 DR
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.036;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
 |||||
 Db 1 EEVVPXGSDYS 11
 |||||

Search completed: June 3, 2004, 11:48:24
 Job time : 46.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	36	66.7	181	4	US-09-134-000C-4848
2	36	66.7	341	3	US-08-853-948B-4
3	36	66.7	348	3	US-08-853-948B-5
4	36	66.7	368	4	US-09-697-367-24
5	36	66.7	1049	4	US-09-394-272-10
6	36	66.7	1068	2	US-08-429-054A-11
7	36	66.7	1068	2	US-08-718-777-7
8	36	66.7	1068	3	US-09-051-341-7
9	36	66.7	1068	4	US-09-394-272-8
10	36	66.7	1081	4	US-09-394-272-4
11	36	66.7	1083	4	US-09-394-272-11
12	36	66.7	1084	4	US-09-394-272-9
13	34	63.0	140	3	US-08-569-147-76
14	34	63.0	140	3	US-08-569-147-82
15	34	63.0	1065	4	US-09-252-991A-31637
16	34	63.0	3472	4	US-09-408-020-4
17	33	61.1	59	4	US-08-963-851-14
18	33	61.1	237	4	US-09-540-236-2677
19	33	61.1	303	4	US-09-134-000C-4318
20	33	61.1	378	1	US-08-070-165F-8
21	33	61.1	378	2	US-08-885-418-8
22	33	61.1	473	4	US-09-252-991A-26805
23	33	61.1	765	4	US-09-819-989-4
24	33	61.1	765	4	US-10-273-992-4
25	33	61.1	801	3	US-09-393-630-6
26	33	61.1	811	4	US-09-819-989-2
27	33	61.1	811	4	US-10-273-992-2

28 33 61.1 883 4 US-09-667-373-4 Sequence 4, Appli
29 32 59.3 65 6 5177197-51 Patent No. 5177197
30 32 59.3 102 2 US-08-580-988A-23 Sequence 23, Appli
31 32 59.3 152 2 US-08-460-694-4 Sequence 4, Appli
32 32 59.3 152 3 US-08-460-744-4 Sequence 4, Appli
33 32 59.3 152 3 US-07-667-711B-4 Sequence 4, Appli
34 32 59.3 173 1 US-08-193-977-7 Sequence 7, Appli
35 32 59.3 189 2 US-08-464-517-21 Sequence 21, Appli
36 32 59.3 189 2 US-08-246-361A-21 Sequence 21, Appli
37 32 59.3 189 3 US-08-463-772-21 Sequence 21, Appli
38 32 59.3 189 5 PCT-US93-05000-21 Sequence 21, Appli
39 32 59.3 231 3 US-08-926-842B-20 Sequence 20, Appli
40 32 59.3 236 2 US-08-464-517-22 Sequence 22, Appli
41 32 59.3 236 2 US-08-246-361A-22 Sequence 22, Appli
42 32 59.3 236 5 PCT-US93-05000-22 Sequence 22, Appli
43 32 59.3 236 5 PCT-US93-05000-22 Sequence 22, Appli
44 32 59.3 240 3 US-08-926-842B-21 Sequence 21, Appli
45 32 59.3 241 4 US-09-107-532A-4172 Sequence 4172, Ap

ALIGNMENTS

RESULT 1

US-09-134-000C-4848
; Sequence 4848, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4848
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 66.7%; Score 36; DB 4; Length 181;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDY 10
Db 145 EEVVPXSEDY 154

RESULT 2

US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:

; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4

; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURE:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; OTHER INFORMATION: Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe,

; OTHER INFORMATION: Pro, Ser, Thr, Trp, Tyr, or Val
US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 228 VIPPGMDFS 236

RESULT 3

US-08-853-948B-5
; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 234 VIPPGMDFS 242

RESULT 4

US-09-697-367-24
; Sequence 24, Application US/09697367
; Patent No. 6323015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Calimi, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Tarczynski, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: B81166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR FILING DATE: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 217 VIPPGMDFS 225

RESULT 5

US-09-394-272-10
; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 436 VIPPGMDFS 444

RESULT 6

US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN; VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Muserlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000

TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068
TYPE: Amino acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide
US-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPGMDFS 443

RESULT 7
US-08-718-777-7
Sequence 7, Application US/08718777
Patent No. 5981852
GENERAL INFORMATION:
APPLICANT: Van Assche, C.
APPLICANT: Lando, D.
APPLICANT: Bruneau, J. M.
APPLICANT: Voelker, T.
APPLICANT: Gervais, M.
TITLE OF INVENTION: MODIFICATION OF SUCROSE
TITLE OF INVENTION: PHOSPHATE
TITLE OF INVENTION: SYNTHASE IN PLANTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Barbara Rae-Venter
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,777
FILING DATE: NOT YET ASSIGNED
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,471

FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.072.02US
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-777-7

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPGMDFS 443

Db 435 VIPGMDFS 443

RESULT 8

US-09-051-341-7
Sequence 7, Application US/09051341
Patent No. 6124528
GENERAL INFORMATION:
APPLICANT: Shewmaker, C. K.
TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rae-Venter Law Group, P.C.
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,341
FILING DATE: 27-OCT-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17351
FILING DATE: 25-OCT-1996
APPLICATION NUMBER: US 08/549,016
FILING DATE: 27-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/372,200
FILING DATE: 12-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter, Ph.D.,
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.110.02US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-051-341-7

Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
Db 435 VIPGMDFS 443

RESULT 9

US-09-394-272-8
Sequence 8, Application US/09394272
Patent No. 6472588
GENERAL INFORMATION:
APPLICANT: Haigler, A. Scott
APPLICANT: Holaday, Candace H.
TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: 201304/1000
CURRENT APPLICATION NUMBER: US/09/394,272
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 14

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Zea mays
US-09-394-272-8

Query Match          66.7%; Score 36; DB 4; Length 1068;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      435 VIPPGMDFS 443

RESULT 10
US-09-394-272-4
; Sequence 4, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-09-394-272-4

Query Match          66.7%; Score 36; DB 4; Length 1081;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      445 VIPPGMDFS 453

RESULT 11
US-09-394-272-11
; Sequence 11, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-394-272-11

Query Match          66.7%; Score 36; DB 4; Length 1083;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      483 VIPPGMDFS 491

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-9

Query Match          66.7%; Score 36; DB 4; Length 1084;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      453 VIPPGMDFS 461

RESULT 12
US-09-394-272-9
; Sequence 9, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-9

Query Match          66.7%; Score 36; DB 4; Length 1084;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      453 VIPPGMDFS 461

RESULT 13
US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-569-147-76

Query Match          63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 VVPXGMDY 10
Db 122 VVPTGFDY 129

RESULT 14
US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSER: No. 6180377ris, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yatko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

Query Match 63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMDY 10
Db 122 VVPTGFDY 129

RESULT 15
US-09-252-991A-31637
; Sequence 31637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31637
; LENGTH: 1065

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31637
Query Match 63.0%; Score 34; DB 4; Length 1065;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMDYS 11
Db 324 PQGMDYS 330

Search completed: June 3, 2004, 12:03:08
Job time : 11.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

2M protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EGVVXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	96.3	11	12	US-09-909-164-8 Sequence 8, Appl
2	52	96.3	11	12	US-09-909-164-12 Sequence 12, Appl
3	52	96.3	11	12	US-09-909-164-13 Sequence 13, Appl
4	47	87.0	11	12	US-09-909-164-22 Sequence 22, Appl
5	47	87.0	11	12	US-09-909-164-26 Sequence 26, Appl
6	47	87.0	11	12	US-09-909-164-27 Sequence 27, Appl
7	46	85.2	11	12	US-09-909-164-5 Sequence 5, Appl
8	46	85.2	11	12	US-09-909-164-6 Sequence 6, Appl
9	46	85.2	11	12	US-09-909-164-9 Sequence 9, Appl
10	46	85.2	11	12	US-09-909-164-10 Sequence 10, Appl
11	46	85.2	11	12	US-09-909-164-31 Sequence 31, Appl
12	46	85.2	11	12	US-09-909-164-32 Sequence 32, Appl
13	46	85.2	11	12	US-09-909-164-35 Sequence 35, Appl
14	46	85.2	11	12	US-09-909-164-40 Sequence 40, Appl
15	46	85.2	11	12	US-09-909-164-41 Sequence 41, Appl

16	46	85.2	11	12	US-09-909-164-45
17	46	85.2	11	12	US-09-909-164-46
18	46	85.2	11	12	US-09-909-164-47
19	46	85.2	11	12	US-09-909-164-48
20	46	85.2	11	12	US-09-909-164-49
21	46	85.2	11	12	US-09-909-164-50
22	46	85.2	11	12	US-09-909-164-51
23	46	85.2	11	12	US-09-909-164-52
24	45	83.3	11	12	US-09-909-164-7
25	45	83.3	11	12	US-09-909-164-11
26	44	81.5	11	12	US-09-909-164-17
27	44	81.5	11	12	US-09-909-164-18
28	41	75.9	11	12	US-09-909-164-19
29	41	75.9	11	12	US-09-909-164-20
30	41	75.9	11	12	US-09-909-164-23
31	41	75.9	11	12	US-09-909-164-24
32	40	74.1	11	12	US-09-909-164-21
33	40	74.1	11	12	US-09-909-164-25
34	40	74.1	11	12	US-09-909-164-28
35	40	74.1	11	12	US-09-909-164-29
36	40	74.1	11	12	US-09-909-164-33
37	40	74.1	11	12	US-09-909-164-36
38	40	74.1	11	12	US-09-909-164-37
39	40	74.1	11	12	US-09-909-164-43
40	39	72.2	11	12	US-09-909-164-30
41	39	72.2	11	12	US-09-909-164-34
42	39	72.2	11	12	US-09-909-164-38
43	39	72.2	11	12	US-09-909-164-39
44	39	72.2	11	12	US-09-909-164-42
45	39	72.2	11	12	US-09-909-164-44

ALIGNMENTS

RESULT 1

US-09-909-164-8
Sequence 8, Application US/0909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NG-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN0192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)

```
; OTHER INFORMATION: AMIDATION
US-09-909-164-8
Query Match          96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
    |||||
Db 1 BEVVPXGMDYS 11

RESULT 2
US-09-909-164-12
; Sequence 12, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; US-09-909-164-13

Query Match          96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
    |||||
Db 1 BEVVPXGMDYS 11

RESULT 4
US-09-909-164-22
; Sequence 22, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-12

Query Match          96.3%; Score 52; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
    |||||
Db 1 BEVVPXGMDYS 11

RESULT 3
US-09-909-164-13
; Sequence 13, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
```


; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-22

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 5
US-09-909-164-26
; Sequence 26, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-26

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 6
US-09-909-164-27
; Sequence 27, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita

; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-27

Query Match 87.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0097;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGQDYS 11

RESULT 7
US-09-909-164-5
; Sequence 5, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)
 FEATURE:
 NAME/KEY: MOD_RES
 LOCATION: (11)..(11)
 OTHER INFORMATION: AMIDATION
 US-09-909-164-5

Query Match 85.2%; Score 46; DB 12; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.015;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 8

US-09-909-164-6
 Sequence 6, Application US/0909164
 Publication No. US20020068702A1
 GENERAL INFORMATION:
 APPLICANT: Corvas International, Inc.
 APPLICANT: Lim-Wilby, Marguerita
 APPLICANT: Levy, Odile E
 APPLICANT: Bruck, Terence K
 TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
 FILE REFERENCE: IN01192-US
 CURRENT APPLICATION NUMBER: US/09/909,164
 PRIOR FILING DATE: 2003-03-25
 PRIOR APPLICATION NUMBER: 60/220,101
 PRIOR FILING DATE: 2000-07-21
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 6

LENGTH: 11
 TYPE: PRT
 ORGANISM: artificial sequence
 FEATURE:
 OTHER INFORMATION: 11-mer synthesized according to example 1.

NAME/KEY: MOD_RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (9)..(9)

OTHER INFORMATION: D-amino acid

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

US-09-909-164-6

Query Match 85.2%; Score 46; DB 12; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.015;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 9

US-09-909-164-9
 Sequence 9, Application US/0909164
 Publication No. US20020068702A1
 GENERAL INFORMATION:
 APPLICANT: Corvas International, Inc.
 APPLICANT: Lim-Wilby, Marguerita

APPLICANT: Levy, Odile E
 APPLICANT: Bruck, Terence K
 TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
 FILE REFERENCE: IN01192-US
 CURRENT APPLICATION NUMBER: US/09/909,164
 CURRENT FILING DATE: 2003-03-25
 PRIOR APPLICATION NUMBER: 60/220,101
 PRIOR FILING DATE: 2000-07-21
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 9

LENGTH: 11
 TYPE: PRT
 ORGANISM: artificial sequence
 FEATURE:
 OTHER INFORMATION: 11-mer synthesized according to example 1

NAME/KEY: MOD_RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (11)..(11)

OTHER INFORMATION: AMIDATION

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (6)..(6)

OTHER INFORMATION: norvaline-(CO)

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (8)..(8)

OTHER INFORMATION: D-amino acid

US-09-909-164-9

Query Match 85.2%; Score 46; DB 12; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 10

US-09-909-164-10
 Sequence 10, Application US/0909164
 Publication No. US20020068702A1
 GENERAL INFORMATION:
 APPLICANT: Corvas International, Inc.
 APPLICANT: Lim-Wilby, Marguerita
 APPLICANT: Levy, Odile E
 APPLICANT: Bruck, Terence K
 TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
 FILE REFERENCE: IN01192-US
 CURRENT APPLICATION NUMBER: US/09/909,164
 CURRENT FILING DATE: 2003-03-25
 PRIOR APPLICATION NUMBER: 60/220,101
 PRIOR FILING DATE: 2000-07-21
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 10

LENGTH: 11

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: 11-mer synthesized according to example 1

NAME/KEY: MOD_RES

LOCATION: (1)..(1)

OTHER INFORMATION: ACETYLTATION

FEATURE:

NAME/KEY: MOD_RES

LOCATION: (11)..(11)

```

; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
JS-09-909-164-10
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2y 1 EEVVPXGMDYS 11
   ||||| ||
2b 1 EEVVPXGMSYS 11

RESULT 11
JS-09-909-164-31
; Sequence 31, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-32
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
   ||||| ||
DB 1 EEVVPXGTDYS 11

RESULT 13
US-09-909-164-35
; Sequence 35, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 35
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-31
Query Match      85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
   ||||| ||
DB 1 EEVVPXGTDYS 11

RESULT 12
US-09-909-164-32
; Sequence 32, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita

```

OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
US-09-909-164-35

Query Match 85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
| | | | | | | | | |
Db 1 BEVVPXGSDYS 11

RESULT 14
US-09-909-164-40
Sequence 40, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
US-09-909-164-40

Query Match 85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
| | | | | | | | | |
Db 1 BEVVPXGSDYS 11

RESULT 15
US-09-909-164-41
Sequence 41, Application US/09909164

Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 41
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-41

Query Match 85.2%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
| | | | | | | | | |
Db 1 BEVVPXGSDYS 11

Search completed: June 3, 2004, 12:57:16
Job time : 33.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	2 S54619	hypothetical prote
2	38	70.4	363	2 D69551	conserved hypothet
3	36	66.7	102	2 A42452	V1 protein - tobac
4	36	66.7	341	2 S72649	sucrose-phosphate
5	36	66.7	348	2 S72650	sucrose-phosphate
6	36	66.7	460	2 G96764	unknown protein F2
7	36	66.7	1049	2 JG4783	sucrose-phosphate
8	36	66.7	1068	1 JQ1329	sucrose-phosphate
9	36	66.7	1081	2 T09837	sucrose-phosphate
10	36	66.7	1083	2 T04062	sucrose-phosphate
11	36	66.7	1084	2 T04103	sucrose-phosphate
12	35	64.8	425	2 T24111	hypothetical prote
13	35	64.8	433	2 H87660	peptidoglycan-bind
14	35	64.8	440	2 H72784	probable alkaline
15	35	64.8	1150	2 T20173	hypothetical prote
16	35	64.8	1474	2 F69009	probable membrane
17	35	64.8	2747	2 B49132	fat facets (fat) s
18	34	63.0	99	2 S00210	plastocyanin b - L
19	34	63.0	155	2 S38255	plastocyanin precu
20	34	63.0	168	2 S58208	plastocyanin b pre
21	34	63.0	290	2 AG3104	6-O-methylguanine-
22	34	63.0	290	2 D98182	Os-methylguanine-D
23	34	63.0	296	2 F72745	hypothetical prote
24	34	63.0	357	1 G69290	probable hexosyltr
25	34	63.0	366	2 G69350	L-lactate dehydrog
26	34	63.0	565	2 E86665	ABC transporter At
27	34	63.0	566	2 A70164	phenylalanine-tRNA
28	34	63.0	587	2 F81138	succinate dehydrog
29	34	63.0	906	2 T48898	disease resistance

30 34 63.0 908 2 T48899 disease resistance
31 34 63.0 1062 2 E83335 RND multidrug effl
32 34 63.0 1062 2 T30830 hypothetical prote
33 34 63.0 3472 2 T31308 hypothetical 367K
34 33 61.1 97 2 A99427 partial transposas
35 33 61.1 128 2 A90471 hypothetical prote
36 33 61.1 172 2 S27021 fibroblast growth
37 33 61.1 184 2 E90335 hypothetical prote
38 33 61.1 225 2 S57310 hypothetical prote
39 33 61.1 247 2 A96001 conserved hypothet
40 33 61.1 257 2 A96546 unknown protein [i
41 33 61.1 262 2 F90298 transposase ISC105
42 33 61.1 267 2 C90307 transposase ISC105
43 33 61.1 276 2 C64417 hypothetical prote
44 33 61.1 283 2 G83055 pantoate-beta-ala
45 33 61.1 299 2 E90487 transposase ISC105

ALIGNMENTS

RESULT 1

S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S54619; S66879
R:de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54617
A:Accession: S54619
A:Molecule type: DNA
A:Residues: 1-156 <DEH>
A:Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123
R:de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66877
A:Accession: S66879
A:Molecule type: DNA
A:Residues: 1-156 <DEW>
A:Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA99201.1; PID:G1420111; MIPS:YOR01
A:Experimental source: strain S288C
C:Genetics:
A:Cross-references: SGD:S0005539
A:Map position: 15R
C:Superfamily: hypothetical protein YOR013w
Query Match 74.1%; Score 40; DB 2; Length 156;
Best Local Similarity 77.8%; Pred. No. 1;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMDY 10

Db 50 EVVPXGMDY 58

RESULT 2

D69551
conserved hypothetical protein AF2411 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: D69551
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 350, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69551
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA
 A/Residues: 1-363 <KLE>
 A/Cross-references: GB:AE001109; GB:AE000782; NID:G2689432; PIDN:AA91255.1; PID:G265068

Query Match 70.4%; Score 38; DB 2; Length 363;
 Best Local Similarity 54.5%; Pred. No. 6.8;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGMDYS 11
 Db 120 ENIVPGIDFS 130

RESULT 3
 A42452
 VI protein - tobacco yellow dwarf virus (strain Australia)
 C/Species: tobacco yellow dwarf virus
 C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 C/Accession: A42452
 R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
 Viology 187, 633-642, 1992
 A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
 A/Reference number: A42452; MUID:92188538; PMID:1546458
 A/Accession: A42452
 A/Molecule type: DNA
 A/Residues: 1-102 <MOR>
 A/Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 66.7%; Score 36; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 4.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMDYS 11
 Db 7 QVVPFGINYS 16

RESULT 4
 S72649
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Citrus unshiu (fragment)
 C/Species: Citrus unshiu
 C/Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
 C/Accession: S72649
 R/Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
 Mol. Gen. Genet. 252, 346-351, 1996
 A/Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
 A/Reference number: S72648; MUID:96439842; PMID:8842155
 A/Accession: S72649
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-341 <KOM>
 A/Cross-references: EMBL:AB006319; NID:G2588891; PIDN:BAA23215.1; PID:G2588892
 A/Experimental source: fruit, cv. Miyagawa-Wase
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
 C/Genetics:
 A/Gene: SP82
 C/Function:
 A/Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:1-341/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 341;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 Db 228 VIPPGMDFS 236

RESULT 5
 JC4783
 sucrose-phosphate synthase (EC 2.4.1.14) - rice
 C/Species: Oryza sativa (rice)

S72650
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 3 - Citrus unshiu (fragment)
 C/Species: Citrus unshiu
 C/Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
 C/Accession: S72650
 R/Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
 Mol. Gen. Genet. 252, 346-351, 1996
 A/Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synth
 A/Reference number: S72648; MUID:96439842; PMID:8842155
 A/Accession: S72650
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-348 <KOM>
 A/Cross-references: EMBL:AB006660; NID:G2351059; PIDN:BAA22071.1; PID:G2351060
 A/Experimental source: fruit, cv. Miyagawa-Wase
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
 C/Genetics:
 A/Gene: SP83
 C/Function:
 A/Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructos
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:1-348/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 348;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 Db 234 VIPPGMDFS 242

RESULT 6
 G96764
 unknown protein P25P22.17 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: G96764
 R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: G96764
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-460 <STO>
 A/Cross-references: GB:AE005173; NID:G6692750; PIDN:AAF24856.1; GSPDB:GN00141
 C/Genetics:
 A/Gene: F25P22.17
 A/Map position: 1

Query Match 66.7%; Score 36; DB 2; Length 460;
 Best Local Similarity 70.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10
 Db 218 EEDVPSAMDY 227

RESULT 7
 JC4783
 sucrose-phosphate synthase (EC 2.4.1.14) - rice
 C/Species: Oryza sativa (rice)

```

C>Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
C/Accession: J04783
R;Valdez-Alarcon, J.J.; Ferrando, M.; Salermo, G.; Jimenez-Moralla, B.; Herrera-Estrella
Gene 170, 217-222, 1996
A/Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
A/Reference number: J04783; MUID:95235138; PMID:8666248
A/Accession: J04783
A/Molecule type: mRNA
A/Residues: 1-1049 <VAL>
A/Cross-references: GB:U33175; NID:G1449931; PIDN:AA049379.1; PID:9988270
A/Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrosephosphate-UDPglucosylt
C/Comment: This enzyme catalyzes the formation of sucrose-phosphate form UDP-glucose and
C/Genetics:
A/Genes: Sps1
A/Introns: 24/1, 103/3; 183/3; 205/3; 435/3; 475/3; 519/3; 578/3; 596/3; 617/3; 931/3; 9
C/Function:
A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
A/Pathway: sucrose biosynthesis
A/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C/Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:180-663/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 436 VIPPGMDFS 444

RESULT 8
JQ1329
sucrose-phosphate synthase (EC 2.4.1.14) - maize
C/Species: Zea mays (maize)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: JQ1329; PQ0260
R;Worrell, A.C.; Bruneau, J.M.; Summerfelt, K.; Boersig, M.; Voelker, T.A.
Plant Cell 3, 1121-1130, 1991
A/Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
A/Reference number: JQ1329; MUID:92338837; PMID:1840396
A/Accession: JQ1329
A/Molecule type: mRNA
A/Residues: 1-1068 <WOR>
A/Cross-references: GB:M97550; NID:G168625; PIDN:AAA33513.1; PID:G168626
A/Accession: PQ0260
A/Molecule type: protein
A/Residues: 71-74; 206-212; 471-481; 872-892 <WOR1>
C/Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
C/Function:
A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
A/Pathway: sucrose biosynthesis
A/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C/Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:178-666/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 435 VIPPGMDFS 443

RESULT 9
T09837
sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantagineum
C/Species: Craterostigma plantagineum
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C/Accession: T09837
R;Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

Plant Physiol. 115, 113-121, 1997
A/Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to suga
A/Reference number: Z16874; MUID:97451773; PMID:9306894
A/Accession: T09837
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1081 <ING>
A/Cross-references: EMBL:Y11795; NID:G2190349; PIDN:CAA72491.1; PID:G2190350
A/Experimental source: ABA-treated callus
C/Genetics:
A/Genes: sps2
C/Function:
A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
A/Pathway: sucrose biosynthesis
A/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C/Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSS>

Query Match 66.7%; Score 36; DB 2; Length 1081;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 445 VIPPGMDFS 453

RESULT 10
T04062
sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
C/Accession: T04062
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A/Reference number: Z15184
A/Accession: T04062
A/Molecule type: DNA
A/Residues: 1-1083 <BEV>
A/Cross-references: EMBL:AL049487
A/Experimental source: cultivar Columbia; BAC clone F28M11
C/Genetics:
A/Map position: 4
A/Introns: 86/3; 116/3; 255/3; 322/2; 482/3; 526/3; 570/3; 629/3; 647/3; 668/3; 949/3; 9
A/Note: F28M11.40
A/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
F:230-714/Domain: sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 483 VIPPGMDFS 491

RESULT 11
T04103
sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
C/Species: Oryza sativa (rice)
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C/Accession: T04103
R;Sakamoto, M.; Satozawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura, T.
Plant Sci. 112, 207-217, 1995
A/Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene that
A/Reference number: Z15212
A/Accession: T04103
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1084 <SAK>
A/Cross-references: EMBL:D45890; PIDN:BA08304.1
A/Experimental source: subsp. Japonica

```

```
C;Genetics:
A;Gene: Spe1
A;Map position: 1
A;Introns: 120/3; 200/2; 221/3; 452/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
C;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C;Keywords: glycosyltransferase; hexosyltransferase
F;196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1084;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 453 VLPFGVDF 461

RESULT 12
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111
A;Reference number: Z19842
A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Accession: T24111
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN: CAB03241.1; GSPDB:GN000023; CESP:R10D12.10
A;Experimental source: clone R10D12
C;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
DB 335 EQIVFGLQY 344

RESULT 13
H87660
peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: H87660
A;Reference number: Z19842
A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Accession: T24111
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN: CAB03241.1; GSPDB:GN000023; CESP:R10D12.10
A;Experimental source: clone R10D12
C;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
DB 335 EQIVFGLQY 344

RESULT 14
H72784
probable alkaline proteinase APE0263 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: H72784
A;Reference number: Z19842
A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Accession: T24111
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN: CAB03241.1; GSPDB:GN000023; CESP:R10D12.10
A;Experimental source: clone R10D12
C;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 440;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMDY 10
DB 120 EVLPWGVY 128

RESULT 15
T20173
hypothetical protein CS3A5.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T20173; T23857
A;Reference number: Z19842
A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Accession: T20173
A;Molecule type: DNA
A;Residues: 1-1150 <WIL>
A;Cross-references: EMBL:Z81486; PIDN: CAB03994.1; GSPDB:GN000023; CESP:CS3A5.2
A;Experimental source: clone CS3A5
A;Map position: 5
A;Introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2; 6

Query Match 64.8%; Score 35; DB 2; Length 1150;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 562 VLPFGVDF 570

Search completed: June 3, 2004, 12:00:01
Job time : 9 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 BEVVPXGMYS 11

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	70.4	1058	1	CARB_FUSNN
2	36	66.7	102	1	Y1LK_TYDVA
3	36	66.7	1049	1	SPS_ORYSA
4	36	66.7	1068	1	SPS_MAIZE
5	36	66.7	1081	1	SPS2_CRAPL
6	35	64.8	2778	1	PAP_DROME
7	34.5	63.9	748	1	KHL1_HUMAN
8	34	63.0	154	1	PLAS_ORYSA
9	34	63.0	155	1	PLAS_HORVU
10	34	63.0	168	1	PLAT_PORNI
11	34	63.0	566	1	SVFB_BORBU
12	34	63.0	908	1	RSL4_ARATH
13	34	63.0	908	1	RPB8_ARATH
14	34	63.0	910	1	RPB8_ARATH
15	33	61.1	276	1	Y339_METJA
16	33	61.1	283	1	PANC_PSEAE
17	33	61.1	394	1	HMPA_VIBCH
18	33	61.1	421	1	ACDM_RAT
19	33	61.1	421	1	ECB2_HALEL
20	33	61.1	423	1	ECB1_HALEL
21	33	61.1	787	1	ECE2_MOUSE
22	33	61.1	801	1	FER3_MOUSE
23	33	61.1	806	1	CEK2_CHICK
24	33	61.1	877	1	SULH_SCHPO
25	33	61.1	982	1	ENV_SFV3L
26	33	61.1	1401	1	RQC_VIBCH
27	33	61.1	2717	1	ZEP1_HUMAN
28	32.5	60.2	472	1	ET2A_XENLA
29	32	59.3	97	1	PLAS_DAUCA
30	32	59.3	150	1	YGC1_PSEPK
31	32	59.3	165	1	TPX_LISIN
32	32	59.3	165	1	TPX_LISMO
33	32	59.3	175	1	HES3_RAT

RESULT 1
CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11899109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fongstein M., Kyripides N., Overbeek R.
RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -|- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -|- PATHWAY: Arginine biosynthesis.
CC -|- PATHWAY: Pyrimidine biosynthesis; first step.
CC -|- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -|- SIMILARITY: Belongs to the carb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A010554; A0194625.1; ALT_INIT.
DR HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CFSase_L_chain; 2.
DR Pfam; PF02786; CFSase_L_D2; 2.
DR Pfam; PF02787; CFSase_L_D3; 1.

P08203 escherichia
P06190 salmonella
Q9wzr1 thermotoga
Q04827 rattus norv
P30279 homo sapien
P30280 mus musculu
Q90459 brachydanio
P50755 xenopus lae
P49706 gallus gall
P53782 xenopus lae
P55169 gallus gall
P30281 homo sapien

ALIGNMENTS

```

DR Pfam; PF02142; MGS; 1.
DR PRINTS; PRO0098; CPSASE.
DR TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
FT SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;
Query Match 70.4%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 9.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGWDYS 11
|:|:|:|:|:|
Db 190 EIVFNGLYNS 199

RESULT 2
Y1LK TYDVA
ID Y1LK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN VI.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
[1]
SEQUENCE FROM N.A.
MEDLINE=92188538; PubMed=1546458;
Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
"The nucleotide sequence of the infectious cloned DNA component of
Tobacco yellow dwarf virus reveals features of geminiviruses
infecting monocotyledonous plants.";
Virology 187:633-642(1992).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M81103; AAA47947.1; -.
PIR; A42452; A42452.
InterPro; IPR002621; Gemini_mov.
Pfam; PF01708; Gemini_mov; 1.
HYPOTHETICAL protein.
KW SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55567 CRC64;
Query Match 66.7%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGWDYS 11
|:|:|:|:|:|

```

```

Db 7 QVPSGINYS 16

RESULT 3
SPS ORYSA
ID SPS ORYSA STANDARD; PRT; 1049 AA.
AC Q43802;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14)
(UDP-glucose-fructose-phosphate glucosyltransferase).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX NCBI_TaxID=4530;
[1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR36; TISSUE=Leaf;
RX MEDLINE=96235138; PubMed=8666248;
RX Valdez-Alarcon J.J., Ferrando M., Jimenez-Morales B.,
Herrera-Estrella L.;
"Characterization of a rice sucrose-phosphate synthase-encoding
Gene.";
Gene 170:217-222(1996).
CC -!- FUNCTION: Involved in the regulation of carbon partitioning in the
leaves of plants. May regulate the synthesis of sucrose and
therefore play a major role as a limiting factor in the export of
photoassimilates out of the leaf.
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
sucrose 6-phosphate.
CC -!- ENZYME REGULATION: Activity regulated by phosphorylation and
moderated by concentration of metabolites and light.
CC -!- PATHWAY: Sucrose synthesis.
CC -!- SUBUNIT: Homodimer or homotetramer (By similarity).
CC -!- PTM: Phosphorylated. However, phosphorylation is not essential for
enzyme function (By similarity).
CC -!- SIMILARITY: Belongs to the glucosyltransferase family 1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U33175; AAC49379.1; -.
PIR; JC4783; JC4783.
Gramene; Q43802;
InterPro; IPR001296; Glyco_transf_1.
Pfam; PF00534; Glycosyltransferase; Phosphorylation.
KW Transferase; Glycosyltransferase; Phosphorylation.
FT DOMAIN 22 29 POLY-GLY.
FT DOMAIN 695 698 POLY-GLU.
FT DOMAIN 775 779 POLY-ARG.
FT SEQUENCE 1049 AA; 116455 MW; ED862E2819AA4B04 CRC64;
Query Match 66.7%; Score 36; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGWDYS 11
|:|:|:|:|:|
Db 436 VIPPGWDFS 444

RESULT 4
SPS MAIZE
ID SPS MAIZE STANDARD; PRT; 1068 AA.
AC P31327;
DT 01-JUL-1993 (Rel. 26, Created)

```

01-JUN-1993 (Rel. 26, Last sequence update)
 01-JUN-1994 (Rel. 29, Last annotation update)
 Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase).
 SPS.
 Zea mays (Maize).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 PACAD clade; Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID=4577;
 [1]
 SEQUENCE FROM N.A., AND SEQUENCE OF 71-74; 206-212; 471-481 AND
 872-892.
 STRAIN=CV, PIONEER 3184; TISSUE=leaf;
 MEDLINE=92338837; Pubmed=1840396;
 Worrell A.C., Brunau J.-W., Summerfelt K., Boersig M., Voelker T.A.;
 "Expression of a maize sucrose phosphate synthase in tomato alters
 leaf carbohydrate partitioning";
 Plant Cell 3:1121-1130(1991).
 -!- FUNCTION: Involved in the regulation of carbon partitioning in the
 leaves of plants. May regulate the synthesis of sucrose and
 therefore play a major role as a limiting factor in the export of
 photoassimilates out of the leaf.
 -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 sucrose 6-phosphate.
 -!- ENZYME REGULATION: Activity regulated by phosphorylation and
 moderated by concentration of metabolites and light.
 -!- PATHWAY: Sucrose synthesis.
 -!- SUBUNIT: Homodimer or homotetramer.
 -!- DEVELOPMENTAL STAGE: Germinating seeds or mature leaves.
 -!- PM: Phosphorylated. However, phosphorylation is not essential for
 enzyme function.
 -!- SIMILARITY: Belongs to the glucosyltransferase family 1.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL; M97550; AAA33513.1; --
 F1R; JQ1329; JQ1329.
 MaizeDB; 25294; --
 InterPro; IPR001296; Glyco.trans.1.
 Pfam; PF00534; Glycos.trans.1; 1.
 Transferase; Glycosyltransferase; Phosphorylation.
 DOMAIN 25 31 POLY-GLY.
 SEQUENCE 1068 AA; 118575 MW; 074679B5E9A1D282 CRC64;

 Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 65.7%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 3 VVPXGMDYS 11
 |:|:|:|:|
 435 VPPGMDFS 443

 RESULT 5
 SP2 CRAPL STANDARD; PRT; 1081 AA.
 ID SP2 CRAPL
 AC 004933;
 15-DEC-1998 (Rel. 37, Created)
 15-DEC-1998 (Rel. 37, Last sequence update)
 15-DEC-1998 (Rel. 37, Last annotation update)
 Sucrose-phosphate synthase 2 (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase 2).
 SPS2
 Craterostigma plantagineum.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

Lamids; Lamiales; Lamiales incertae sedis; Lindernieae;
 Craterostigma.
 NCBI_TaxID=4153;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=97451773; Pubmed=9306694;
 Ingram J., Chandler J.W., Gallagher L., Salamini F., Bartels D.;
 "Analysis of cDNA clones encoding sucrose-phosphate synthase in
 relation to sugar interconversions associated with dehydration in the
 resurrection plant Craterostigma plantagineum Hochst.";
 Plant Physiol. 115:113-121(1997).
 -!- FUNCTION: Involved in the regulation of carbon partitioning in the
 leaves of plants. May regulate the synthesis of sucrose and
 therefore play a major role as a limiting factor in the export of
 photoassimilates out of the leaf.
 -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 sucrose 6-phosphate.
 -!- ENZYME REGULATION: Activity regulated by phosphorylation and
 moderated by concentration of metabolites and light.
 -!- PATHWAY: Sucrose synthesis.
 -!- SUBUNIT: Homodimer or homotetramer (By similarity).
 -!- PM: Phosphorylated. However, phosphorylation is not essential for
 enzyme function (By similarity).
 -!- SIMILARITY: Belongs to the glucosyltransferase family 1.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL; Y11795; CAA72491.1; --
 F1R; T09837; T09837.
 InterPro; IPR001296; Glyco.trans.1.
 Pfam; PF00534; Glycos.trans.1; 1.
 Transferase; Glycosyltransferase; Phosphorylation; Multigene family.
 DOMAIN 245 248 POLY-SER.
 DOMAIN 256 264 POLY-GLU.
 DOMAIN 787 790 POLY-ARG.
 SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;

 Query Match 66.7%; Score 36; DB 1; Length 1081;
 Best Local Similarity 65.7%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 3 VVPXGMDYS 11
 |:|:|:|:|
 445 VPPGMDFS 453

 RESULT 6
 FAF DROME STANDARD; PRT; 2778 AA.
 ID FAF DROME
 AC P55824; Q9V9T6; Q9V0Z7;
 01-NOV-1997 (Rel. 35, Created)
 28-FEB-2003 (Rel. 41, Last sequence update)
 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
 (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
 FAF) (Deubiquitinating enzyme FAF) (Fat facets' protein).
 GN FAF OR BCDNA:LD22582 OR CG1945.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 [1]
 SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
 TISSUE=Eye imaginal disk;
 MEDLINE=93202020; Pubmed=1295747;
 Fischer-Vize J.A., Rubin G.M., Lehmann R.;

RT "The fat facets gene is required for Drosophila eye and embryo
 RL development.";
 RL Development 116:985-1000(1992).

[2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.P., Agbayani A., An H.-J., Andrews-Framknoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostier C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.A., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy L., Murphy L., Muzny D.N., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sinden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Virsikas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Wungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.F.,
 RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 FT "Annotation of the Drosophila melanogaster euchromatic genome: a
 FT systematic review";
 FT Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RL [4]
 RP SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).
 RN STRAIN=Berkley;
 RX MEDLINE=20196012; PubMed=10731138;
 RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
 RA Stapleton M., Harvey D.A.;
 RT "A Drosophila complementary DNA resource.";
 RL Science 287:2222-2224(2000).
 CC -1- FUNCTION: Required for eye and embryo development, and plays a
 CC role in compound eye assembly and oogenesis respectively. In the
 CC larval eye disks, cells outside the assembling facets require this
 CC protein for short-range cell interactions that prevent the mystery
 CC cells from becoming photoreceptors. It is also required for
 CC nuclear migration and cellularization in early embryogenesis and
 CC could play a role in pole cell determination, development or

function.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=P55824-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P55824-2; Sequence=VSP_005270;
 CC Name=3;
 CC IsoId=P55824-3; Sequence=VSP_005269;
 CC -1- TISSUE SPECIFICITY: Eye disks and ovaries.
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -1- SIMILARITY: Belongs to peptidase family C19.
 CC This SWISS-PROT entry is copy-right. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04959; AAF01345.1; -;
 CC EMBL; L04958; AAF01346.1; -;
 CC EMBL; L04960; AAF01347.1; -;
 CC EMBL; L04960; AAF01348.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AF145677; AAD4291.1; -;
 CC MEROPS; C19.007; -;
 CC FLYBase; FBgn005632; faf.
 CC GO; GO:0005737; C:cyttoplasm; IDA.
 CC GO; GO:0007349; P:cellularization; IMP.
 CC GO; GO:0007993; P:embryonic morphogenesis; IMP.
 CC GO; GO:0007456; P:eye morphogenesis (sensu Drosophila); IMP.
 CC GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. .; IMP.
 CC GO; GO:0007097; P:nuclear migration; IMP.
 CC GO; GO:0016579; P:protein denaturation; IDA.
 CC GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IGI.
 CC InterPro; IPR001394; Peptidase_C19.
 CC Pfam; PF00443; UCH; 1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS00973; UCH_2_3; 1.
 CC PROSITE; PS02335; UCH_2_3; 1.
 CC Ubl conjugation pathway; Hydrolase; Thiol protease;
 CC Developmental protein; Vision; Alternative splicing.
 CC ACT SITE 1677 1677 BY SIMILARITY.
 CC ACT SITE 1978 1978 BY SIMILARITY.
 CC ACT SITE 1986 1986 BY SIMILARITY.
 CC VARSPLIC 2705 2778
 CC
 CC VARSPLIC 2742 2778
 CC
 CC FT CONFLICT 234 234
 CC FT CONFLICT 2725 2725
 CC FT SEQUENCE 2778 AA; 311139 MW; FFB90438EA53A02B CRC64;
 CC
 CC Query Match 64.8%; Score 35; DB 1; Length 2778;
 CC Best Local Similarity 54.5%; Pred. No. 1.1e+02;
 CC Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 CC
 CC QY 1 BEVFXGMDYS 11
 CC Db 1394 EVIVPDGQDFS 1404
 CC
 CC RESULT 7

KLHL1 HUMAN
ID Q9NR64; Q9H4X4; STANDARD; PRT; 748 AA.
AC Q9NR64; Q9H4X4; Q9NR65; Q9P238;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kelch-like protein 1.
EN KLHL1 OR KIAA1490.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347694; PubMed=10898605;
RA Koob M.D., Nemes J.P., Benzow K.A.;
RT "The SC48 transcript is an antisense RNA to a brain-specific
transcript encoding a novel actin-binding protein (KLHL1).";
RL Hum. Mol. Genet. 9:1543-1551(2000).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
[3]
RP SEQUENCE OF 179-409 FROM N.A.
RA Kay M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in organizing the actin cytoskeleton of
the brain cells.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highly expressed in brain.
CC -!- SIMILARITY: Contains 1 B1B/POZ domain.
CC -!- SIMILARITY: Contains 6 Kelch repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF252283; AAF81719.1; -.
DR EMBL; AF252279; AAF81716.1; -.
DR EMBL; AB040923; BA96014.1; ALT_INIT.
DR EMBL; AL353738; CAC16128.1; -.
DR Genbank; HGNC:6352; KLHL1.
DR MIM; 605332; -.
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0003779; F:actin binding; NAS.
DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; NAS.
DR InterPro; IPR000210; B1B_POZ.
DR InterPro; IPR008652; Kelch_rep.
DR Pfam; PF00651; B1B; 1.
DR Pfam; PF01344; Kelch; 6.
DR SMART; SM00225; B1B; 1.
DR SMART; SM00612; Kelch; 6.
DR PROSITE; PS50097; B1B; 1.
KW Cytoskeleton; Actin-binding; Kelch repeat; Repeat.
FT DOMAIN 43 88
FT DOMAIN 212 279
FT REPEAT 460 506
FT REPEAT 507 553
FT REPEAT 555 600
FT REPEAT 601 647
FT REPEAT 649 700
FT REPEAT 701 747
FT REPEAT 748 AA; 82680 MW; C11C43D82829FF9 CRC64;
SQ SEQUENCE

Query Match 63.9%; Score 34.5; DB 1; Length 748;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 1 EEVVPXGMDY 10
DB 127 EEVVP-GMDF 135
RESULT 8
PLAS ORYSA STANDARD; PRT; 154 AA.
ID P20423; Q9SBB8;
AC P20423; Q9SBB8; (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
EN PETE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
NCBI TaxID=4530;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ilpoom; TISSUE=Leaf;
RA Lee J.-S.;
RT "Molecular cloning and characterization of plastocyanin precursor in
rice.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 58-154.
RX STRAIN=cv. Japonica;
RX MEDLINE=89386623; PubMed=2780537;
RA Yano H., Kano M., Tsugita A., Aso K., Nozu Y.;
RT "The amino acid sequence of plastocyanin from rice (Oryza sativa,
subspecies japonica).";
RL Protein Seq. Data Anal. 2:385-389(1989).
CC -!- FUNCTION: Participates in electron transfer between P700 and the
cytochrome b6-f complex in photosystem I.
CC -!- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid
membrane surface in chloroplasts.
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF093636; AAC78108.1; -.
DR HSP; P00289; 2PCF.
DR Gramene; P20423; -.
DR InterPro; IPR000923; BlueCu_1.
DR InterPro; IPR001235; Copper_blue.
DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR ProDom; PD001235; Copper_blue; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
Transit peptide.
FT TRANSIT 1 57
FT CHAIN 58 154
FT DOMAIN 58 154
FT METAL 94 94
FT METAL 139 139
FT METAL 142 142
FT METAL 147 147
FT SEQUENCE 154 AA; 15577 MW; E45725D25B5F400D CRC64;
SQ

Query Match 63.0%; Score 34; DB 1; Length 154;
 Best Local Similarity 54.5%; Pred. NO. 8.7;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EGVVXGMDYS 11
 : : : : :
 DB 100 EDAVPSGVDVS 110

RESULT 9

ID	PLAS HORVU	STANDARD;	PRT;	155 AA.
AC	P08248;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Plastocyanin, chloroplast precursor.			
GN	PETE.			
OS	Hordeum vulgare (Barley).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;			
OC	Triticeae; Hordeum.			
OX	NCBI_TaxID=4513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Bomi;			
RA	Nielsen O.S., Gausing K.;			
RT	"The precursor of barley plastocyanin: sequence of cDNA clones and			
RT	gene expression in different tissues.";			
RL	FEBS Lett. 225:159-162(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. NK 1558;			
RA	MEDLINE=94039081; PubMed=8223592;			
RT	Nielsen P., Gausing K.;			
RT	"In vitro binding of nuclear proteins to the barley plastocyanin gene			
RT	promoter region.";			
RL	Eur. J. Biochem. 217:97-104(1993).			
CC	-1- FUNCTION: Participates in electron transfer between P700 and the			
CC	cytochrome b6-f complex in photosystem I.			
CC	-1- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid			
CC	membrane surface in chloroplasts.			
CC	-1- SIMILARITY: Contains 1 plastocyanin-like domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; Y00704; CAA68696.1; -			
DR	EMBL; Z28347; CAA82201.1; -			
DR	PIR; S38255; S38255.			
DR	HSSP; P00289; 2PCF.			
DR	InterPro; IPR000923; BlueCu 1.			
DR	InterPro; IPR001235; Copper_blue.			
DR	InterPro; IPR008972; Cupredoxin.			
DR	Pfam; PF00127; copper-bind; 1.			
DR	PRINTS; PR00156; COPPERBLUE.			
DR	ProDom; PD001235; Copper_blue; 1.			
DR	PROSITE; PS00196; COPPER_BLUE; 1.			
KW	Chloroplast; Electron transport; Copper; Thylakoid; Membrane;			
KW	Transit peptide.			
FT	TRANSIT 1 58 CHLOROPLAST.			
FT	CHAIN 59 155 PLASTOCYANIN.			
FT	DOMAIN 59 155 PLASTOCYANIN-LIKE.			
FT	METAL 95 95 COPPER (BY SIMILARITY).			
FT	METAL 140 140 COPPER (BY SIMILARITY).			
FT	METAL 143 143 COPPER (BY SIMILARITY).			
FT	METAL 148 148 COPPER (BY SIMILARITY).			

FT VARIANT 120 120 T -> N (IN CV. NK 1558).
 SQ SEQUENCE 155 AA; 15709 MW; DAA7EABESF6F4F91 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 155;
 Best Local Similarity 54.5%; Pred. NO. 8.7;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EGVVXGMDYS 11
 : : : : :
 DB 101 EDAVPSGVDVS 111

RESULT 10

ID	PLAT POPNI	STANDARD;	PRT;	168 AA.
AC	P11970;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Plastocyanin B, chloroplast precursor.			
GN	PETE.			
OS	Populus nigra (Lombardy poplar).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid 1; Malpighiales; Salicaceae; Salicaceae; Populus.			
OX	NCBI_TaxID=3691;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Italica;			
RA	STRAIN=cv. Italica; TISSUE=Leaf;			
RA	Dimitrov M.I., Egorov C.A., Donchev A.A., Atanasov B.P.;			
RT	"Complete amino acid sequence of poplar plastocyanin B.";			
RL	FEBS Lett. 226:17-22(1987).			
CC	-1- FUNCTION: Participates in electron transfer between P700 and the			
CC	cytochrome b6-f complex in photosystem I.			
CC	-1- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid			
CC	membrane surface in chloroplasts.			
CC	-1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF			
CC	POPULAR PLASTOCYANINS A AND B.			
CC	-1- SIMILARITY: Contains 1 plastocyanin-like domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; Z50186; CAA90565.1; -			
DR	PIR; S00210; S00210.			
DR	PIR; S58208; S58208.			
DR	HSSP; P00299; 1PLC.			
DR	InterPro; IPR000923; BlueCu 1.			
DR	InterPro; IPR001235; Copper_blue.			
DR	InterPro; IPR008972; Cupredoxin.			
DR	Pfam; PF00127; copper-bind; 1.			
DR	PRINTS; PR00156; COPPERBLUE.			
DR	ProDom; PD001235; Copper_blue; 1.			
DR	PROSITE; PS00196; COPPER_BLUE; 1.			
KW	Chloroplast; Electron transport; Copper; Thylakoid; Membrane;			
KW	Transit peptide; Multigene family.			
FT	TRANSIT 1 69 CHLOROPLAST.			
FT	CHAIN 70 168 PLASTOCYANIN B.			
FT	DOMAIN 70 168 PLASTOCYANIN-LIKE.			
FT	METAL 106 106 COPPER.			
FT	METAL 153 153 COPPER.			
FT	METAL 156 156 COPPER.			
FT	METAL 161 161 COPPER.			

```
SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2038AEEA CRC64;
Query Match 63.0%; Score 34; DB 1; Length 168;
Best Local Similarity 54.5%; Pred. No. 9.5;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

2Y 1 EVVPXGMDYS 11
|:|:|:|:|
3B 112 EDVPSGVDS 122

RESULT 11
SYFB_BORBU
ID SYFB_BORBU STANDARD; PRT; 566 AA.
AC P94283;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PheRS).
DE PHET OR BB0514.
3N Borrelia burgdorferi (lyme disease spirochete).
3S Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
3C NCBI TaxID=139;
3X [1]
3N SEQUENCE FROM N.A.
3R STRAIN-ATCC 35210 / B31;
3C STRAIN-ATCC 35210 / B31;
3A Barbour A.G.; Hinnebusch J.;
3R "Phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and
3T thiodoxin reductase gene of Borrelia burgdorferi."
3L Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
3N [2]
3R SEQUENCE FROM N.A.
3C STRAIN-ATCC 35210 / B31;
3A Frazer C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
3R Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
3R Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
3A Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
3A Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Waldman J.,
3A Uterback T., Wathey L., McDonald L., Artiach P., Bowman C.,
3A Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
3A Smith H.O., Venter J.C.;
3R "Genomic sequence of a Lyme disease spirochaete, Borrelia
3T burgdorferi."
3L Nature 390:580-586(1997).
3C -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA (Phe) = AMP +
3C diphosphate + L-phenylalanyl-tRNA(Phe).
3C -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
3C similarity).
3C -!- SUBCELLULAR LOCATION: Cytoplasmic.
3C -!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
3C family. Subfamily 2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U82978; AAS41019.1; -.
EMBL; AB001153; AAC66870.1; -.
PIR; A70164; A70164.
TIGR; BB0514; -.
HAMAP; MF_00284; -. 1.
InterPro; IPR005147; B5.
InterPro; IPR004531; PheT_arch.
Pfam; PF03484; B5; 1.
TIGRFAM; TIGR00471; PheT_arch; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Complete proteome.

SQ SEQUENCE 566 AA; 65173 MW; 9D48C8B5D6D3B74B CRC64;
Query Match 63.0%; Score 34; DB 1; Length 566;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPXGMDY 10
|:|:|:|:|
DB 169 VPFGMDY 175

RESULT 12
RSL4_ARATH
ID RSL4_ARATH STANDARD; PRT; 908 AA.
AC Q9FJK8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
CN RPP8L4 OR A75G48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
3N [1]
3R SEQUENCE FROM N.A.
3C STRAIN-cv. Columbia;
3X MEDLINE=99087489; PubMed=9872454; Kaneko T., Kotani H., Miyajima N.,
3A Nakamura Y., Sato S., Asamizu E., Narusaka M., Carninci P., Kawai J.,
3R Tabata S.;
3T "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
3L Sequence features of the regions of 1,013,767 bp covered by sixteen
3R physically assigned P1 and TAC clones."
3L DNA Res. 5:237-308(1998).
3N [2]
3R SEQUENCE FROM N.A.
3C STRAIN-cv. Columbia;
3X Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
3R Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
3R Hayashizaki Y., Shinozaki K.;
3T "Arabidopsis thaliana full-length cDNA."
3L Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
3C -!- FUNCTION: Potential disease resistance protein.
3C -!- DOMAIN: The LRR repeats probably act as specificity determinant of
3C pathogen recognition (By similarity).
3C -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
3C RPP8/HRT subfamily.
3C -!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
3C -!- SIMILARITY: Contains 1 NB-ARC domain.
3C -!- DATABASE: NAME=NIB-LRRS;
3C NOTE=Functional and comparative genomics of disease resistance gene
3C homologs;
3C WWW=http://niblr.s.uctavis.edu".
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB015468; BAB10695.1; -.
EMBL; AK117163; BAC41841.1; -.
InterPro; IPR000767; Disease_resist.
InterPro; IPR001611; LRR.
InterPro; IPR002182; NB-ARC.
Pfam; PF00560; LRR; 2.
Pfam; PF00931; NB-ARC; 1.
PRINTS; PR00364; DISEASERESIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
DOMAIN 10 45
LEUCINE-ZIPPER.
```


FT DOMAIN 146 459 NB-ARC.
 FT REPEAT 575 599 LRR 1.
 FT REPEAT 600 623 LRR 2.
 FT REPEAT 842 867 LRR 3.
 FT NP BIND 192 199 ATP (POTENTIAL).
 SQ SEQUENCE 908 AA; 104448 MW; 3111951B17239693 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 54;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPKGMDD 10
 DB 883 EKLVPGEDY 892

RESULT 13
 RPP8 ARATH STANDARD; PRT; 908 AA.
 AC Q9W4J9; Q9W5A1; Q9ZSV3; Q9ZSV4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Disease resistance protein RPP8 (Resistance to *Peronospora parasitica* protein 8).
 GN RPP8 OR HRT OR AT5G43470 OR MWF20.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Arabidopsi Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eukaryota; Viridiplantae; Streptophyta; eudicotyledons; core eudicots; rosids;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Arabidopsi.
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND RPP8-3, AND VARIANTS.
 RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
 RX MEDLINE=99030193; PubMed=9811794;
 RA McDowell J.M., Dhandaaydhay M., Long T.A., Aarts M.G.M., Goff S., Holub E.B., Dang J.L.;
 RA "Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsi";
 RT Plant Cell 10:1861-1874(1998).
 RL [2]
 RN SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
 RC STRAIN=cv. Di-17;
 RX MEDLINE=20271766; PubMed=10810142;
 RA Cooley M.B., Pathirana S., Wu H.-J., Kachroo P., Klessig D.F.;
 RA "Members of the Arabidopsi HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens.";
 RT Plant Cell 12:663-676(2000).
 RL [3]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H., Tabata S.;
 RA "Structural analysis of Arabidopsi thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones.";
 RT DNA Res. 7:31-63(2000).
 RL [4]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RX MEDLINE=23954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Southwick A.M., Wu H.-C., Kim C.-J., Nguyen M., Pham P.K., Cheuk R.F., Karlin-Newmann G., Liu S.-X., Lam B., Sakano H., Wu T., Yu G., Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J., Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y., Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P., Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F., Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M., Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J., Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T., Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y., Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RA "Empirical analysis of transcriptional activity in the Arabidopsi genome.";
 RT Science 302:842-846(2003).
 RL [5]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J., Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J., Hayashizaki Y., Shinozaki K.;
 RA "Arabidopsi thaliana full-length cDNA.";
 RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RL [6]
 RN INTERACTION WITH TIP.
 RP MEDLINE=20496823; PubMed=11041886;
 RX Ren T., Qu F., Morris T.J.;
 RA "HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus.";
 RT Plant Cell 12:1917-1926(2000).
 RL CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard the plant against pathogens that contain an appropriate avirulence protein via an indirect interaction with this avirulence protein. That triggers a defense system including the hypersensitive response, which restricts the pathogen growth. The interaction with TIP (TCV-interacting protein) may be essential for the recognition of the avirulence proteins, and the triggering of the defense response.
 CC CC -1- SUBUNIT: Interacts with the NAC protein TIP.
 CC CC -1- ALTERNATIVE PRODUCTS:
 CC CC Event-Alternative splicing, Named isoforms=2;
 CC CC Name=1;
 CC CC IsoId=Q9W4J9-1; Sequence=Displayed;
 CC CC Name=2;
 CC CC IsoId=Q9W4J9-2; Sequence=VSP_007171, VSP_007172;
 CC CC Notes=Has been shown to exist only in cv. Columbia so far;
 CC CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of pathogen recognition.
 CC CC -1- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and the highly related RPP8 and RPP8A genes present in cv. Landsberg erecta. Such variations probably modify the specificity of pathogen recognition.
 CC CC -1- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8 specifically recognizes the Emco5 avirulence protein from *Peronospora parasitica*, while it is not the case in cv. Di-17, where it confers resistance to Turnip Crinkle Virus upon recognition of the viral capsid protein.
 CC CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family. RPP8/HRT subfamily.
 CC CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC CC -1- DATABASE: NAME=NTB-LRRS;
 CC CC NOTES=Functional and comparative genomics of disease resistance gene homologs;
 CC CC WWW="http://nblrrs.ucdavis.edu".

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF089710; AAC83165.1; -
 EMBL; AF089711; AAC78631.1; -
 EMBL; AF234174; AAF36587.1; -
 EMBL; AB025638; BAA97426.1; -
 EMBL; AY062514; AAL32592.1; -
 EMBL; AK118862; BAC34449.1; -
 InterPro; IPR000767; Disease_resist.


```
RESULT 15
Y939_METJA STANDARD; PRT; 276 AA.
AC Q58349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0939.
GN MJ0939.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073 (1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67537; AB98946.1; -.
DR FIR; C64417; C64417.
DR TIGR; MJ0939; -.
DR InterPro; IPR001163; snRNP Sm.
DR InterPro; IPR005358; UPF0153.
DR Pfam; PF03692; UPF0153; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 276 AA; 33454 MW; 97BD69D392BC8FDF CRC64;
SQ
Query Match 61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred. NO. 26;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EEVVPXGMDS 11
Db 141 EEIENGWERS 151
Search completed: June 3, 2004, 11:49:53
Job time : 4.86667 secs
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	3 Q12479	Q12479 saccharomyc
2	38	70.4	319	16 Q8ESV7	Q8ESV7 oceanobacil
3	38	70.4	363	17 Q30260	Q30260 archaeoglob
4	38	70.4	1044	16 Q8DIH0	Q8DIH0 synectococc
5	36	66.7	149	16 Q822B4	Q822B4 enterococcu
6	36	66.7	341	10 Q22081	Q22081 citrus unsh
7	36	66.7	348	10 Q22096	Q22096 citrus unsh
8	36	66.7	452	10 Q8W568	Q8W568 arabidopsis
9	36	66.7	460	10 Q9C9T7	Q9C9T7 arabidopsis
10	36	66.7	1047	10 Q93782	Q93782 saccharum o
11	36	66.7	1083	10 Q9SN30	Q9SN30 arabidopsis
12	36	66.7	1084	10 Q43010	Q43010 oryza sativ
13	36	66.7	1100	10 Q8S064	Q8S064 eriocheir s
14	35	64.8	219	5 Q9GQ04	Q9GQ04 clostridium
15	35	64.8	253	16 Q8XPA8	Q8XPA8 rhizobium t
16	35	64.8	298	2 Q52367	Q52367 rhizobium t

17	35	64.8	308	16	Q88CC0	Q88CC0 pseudomonas
18	35	64.8	368	2	Q8XL43	Q8XL43 rhizobium e
19	35	64.8	425	5	Q9XVK4	Q9XVK4 caenorhabdi
20	35	64.8	433	16	Q9A382	Q9A382 caulobacter
21	35	64.8	440	17	Q9VFI3	Q9VFI3 aeropyrum p
22	35	64.8	474	16	Q83L32	Q83L32 ehigella fl
23	35	64.8	511	2	Q52680	Q52680 escherichia
24	35	64.8	517	16	Q8XZL5	Q8XZL5 raistonia s
25	35	64.8	595	10	Q94GJ7	Q94GJ7 andrographi
26	35	64.8	745	5	Q95P46	Q95P46 carcinus ma
27	35	64.8	1031	5	Q9U6A3	Q9U6A3 callinectes
28	35	64.8	1150	5	Q17704	Q17704 caenorhabdi
29	35	64.8	1410	2	Q52673	Q52673 escherichia
30	35	64.8	1420	2	Q52666	Q52666 escherichia
31	35	64.8	1474	17	Q27146	Q27146 methanobact
32	35	64.8	1828	16	Q98K29	Q98K29 rhizobium l
33	34.5	63.0	748	4	Q8TEJ7	Q8TEJ7 homo sapien
34	34	63.0	215	16	Q8R9L5	Q8R9L5 thermococ
35	34	63.0	222	16	Q8E5L9	Q8E5L9 streptococc
36	34	63.0	222	16	Q8DZM9	Q8DZM9 streptococc
37	34	63.0	245	16	Q7V6Q4	Q7V6Q4 prochloroco
38	34	63.0	251	9	Q855K6	Q855K6 mycobacteri
39	34	63.0	284	2	Q8GP33	Q8GP33 lactobacill
40	34	63.0	290	16	Q8U7J0	Q8U7J0 agrobacteri
41	34	63.0	296	17	Q9YET8	Q9YET8 aeropyrum p
42	34	63.0	337	16	Q7UF95	Q7UF95 rhodospirell
43	34	63.0	344	16	Q815A7	Q815A7 bacillus ce
44	34	63.0	357	17	Q29920	Q29920 archaeoglob
45	34	63.0	366	17	Q29451	Q29451 archaeoglob

ALIGNMENTS

RESULT 1

Q12479	PRELIMINARY;	PRT;	156 AA.
AC Q12479			
DT 01-NOV-1996 (TEMBLrel. 01, Created)			
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)			
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)			
DE ORF YOR013W.			
GN YOR013W.			
OS Saccharomyces cerevisiae (Baker's yeast).			
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX NCBI_TaxID=4932;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA De haan M., Grivell L.A., Maarse A.C.;			
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA MIPS;			
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RA STRAIN=FY1679;			
RC De haan M., Maarse A.C., Grivell L.A.;			
RN [4]			
RP Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
RN [5]			
RP SEQUENCE FROM N.A.			
RA STRAIN=FY1679;			
RC MEDLINE=94019318; PubMed=8413243;			
RN Dumont M.E., Schlachter J.B., Cardillo T.S., Hayes M.K., Bethlenny G.,			
RA Sherman F.;			
RT "CYC2 encodes a factor involved in mitochondrial import of yeast			
RL cytochrome c.";			
RL Mol. Cell. Biol. 13:6442-6451(1993).			
RN [5]			
RP SEQUENCE FROM N.A.			
RA STRAIN=FY1679;			
RC MEDLINE=94169519; PubMed=7764548;			
RN [5]			

```

RA Lee Y.S., Shimizu J., Yoda K., Yamasaki M.;
RT "Molecular cloning of a gene, DHS1, which complements a drug-
RT hypersensitive mutation of the yeast Saccharomyces cerevisiae.";
RL Biosci. Biotechnol. Biochem. 58:391-395(1994).
DR EMBL; Z74920; CAA59201.1; -.
DR ENBL; X87331; CAA60762.1; -.
DR PIR; S54619; S54619.
DR SGD; S0005539; YOR013W.
SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;

Query Match 74.1%; Score 40; DB 3; Length 156;
Best Local Similarity 77.8%; Pred. No. 3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMDY 10
Db 50 EVVPLGMDY 58
:::|::|::|

RESULT 2
Q8ESV7 PRELIMINARY; PRT; 319 AA.
AC Q8ESV7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical conserved protein.
GN O80509.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE931 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004594; BAC12465.1; -.
DR InterPro; IPR001279; Elactamase-like.
DR Pfam; PF00753; lactamase_B; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDAE4BF13E79E37 CRC64;

Query Match 70.4%; Score 38; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGMDY 10
Db 189 EQVPHGIDY 198
:::|::|::|

RESULT 3
O30260 PRELIMINARY; PRT; 363 AA.
AC O30260;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein AF2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

```

```

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001109; AAB91255.1; -.
DR PIR; D69551; D69551.
DR TIGR; AF2411; -.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; Bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;

Query Match 70.4%; Score 38; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
Db 120 ENVPYGDIFS 130
:::|::|::|

RESULT 4
Q8DIH0 PRELIMINARY; PRT; 1044 AA.
AC Q8DIH0;
DT 01-VAR-2003 (TrEMBLrel. 23, Created)
DT 01-VAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpou S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AF005374; BAC09170.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001036; Acriflavin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 70.4%; Score 38; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
Db 843 BEVLENGIGYS 853
:::|::|::|

```

RC	STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
RA	Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
RR	"Differential expression of three sucrose-phosphate synthase isoforms
RT	during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL	Plant Sci. 140:169-178(1999).
DR	BMBL; AB006319; BAA23215.1; -.
DR	PIR; S72649; S72649.. 1
FT	NON_TER 341 341
FT	NON_TER 341 FT
SQ	SEQUENCE 341 AA; 38136 MW; 61417A69C4560777 CRC64;
Query Match 66.7%; Score 36; DB 10; Length 341;	
Best Local Similarity 66.7%; Pred. No. 50;	
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps	
QY	3 VVPXGMDYS 11
DB	228 VIPFGMDFS 236
RESULT 7	
OZ2096	PRELIMINARY; PRT; 348 AA.
ID OZ2096	
AC OZ2096;	
DT 01-JAN-1998 (TrEMBLrel. 05, Created)	
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)	
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)	
DE Sucrose-phosphate synthase (Fragment).	
GN CITSPS3.	
OS Citrus unshiu (Satsuma orange).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;	
OC eurosoids II; Sapindales; Rutaceae; Citrus.	
OX NCBI_TaxID=55186;	
OX [1]	
RP SEQUENCE FROM N.A.	
RP TISSUE=Juice sacs and segment epidermis;	
RX MEDLINE=96439842; PubMed=842155;	
RA Komatsu A., Takanokura Y., Omura M., Akihama T.;	
RA "Cloning and molecular analysis of cDNA encoding three sucrose	
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu	
RT Marc.).";	
RL Mol. Gen. Genet. 252:346-351(1996).	
RL [2]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Juice sacs and segment epidermis;	
RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;	
RA "Differential expression of three sucrose-phosphate synthase isoforms	
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";	
RL Plant Sci. 140:169-178(1999).	
DR EMBL; AB006660; BAA22071.1; -.	
DR PIR; S72650; S72650.. 1	
FT NON_TER 348 348	
FT NON_TER 348 FT	
SQ SEQUENCE 348 AA; 38556 MW; EE1C21EBA6FF5C5E CRC64;	
Query Match 66.7%; Score 36; DB 10; Length 348;	
Best Local Similarity 66.7%; Pred. No. 51;	
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps	
QY	3 VVPXGMDYS 11
DB	234 VIPFGMDFS 242
RESULT 8	
ID Q8W568	PRELIMINARY; PRT; 452 AA.
AC Q8W568;	
DT 01-MAR-2002 (TrEMBLrel. 20, Created)	
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)	
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE At1g73750/F25P22.17.	

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carrincci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
 RA Carrincci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 DR EMBL; AF419606; AA131938.1; -;
 DR EMBL; AY113044; AA47352.1; -;
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR GO; GO:0004287; F: proteolytic activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estr.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 SQ SEQUENCE 452 AA; 49682 MW; A159955B21742C4A CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 452;
 Best Local Similarity 70.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EVVVPXGMDY 10
 Db 210 EEDVPSAMDY 219
 Query Match 66.7%; Score 36; DB 10; Length 452;
 Best Local Similarity 70.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EVVVPXGMDY 10
 Db 210 EEDVPSAMDY 219

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Miltscher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
 RA Rai C., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana";
 RL Nature 408:816-820(2000).
 DR EMBL; AC012679; AG52073.1; -;
 DR FIR; G96764; G96764.
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR GO; GO:0004287; F: proteolytic activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002471; Pept_S9_AS.
 DR InterPro; IPR000379; Ser_estr.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 460 AA; 50564 MW; E94B27B5C4B249EC CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 460;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EVVVPXGMDY 10
 Db 218 EEDVPSAMDY 227
 Query Match 66.7%; Score 36; DB 10; Length 460;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EVVVPXGMDY 10
 Db 218 EEDVPSAMDY 227

RESULT 10
 P93782 PRELIMINARY; PRT; 1047 AA.
 ID P93782
 AC P93782
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (Fragment).
 GN SOSPS1.
 OS Saccharum officinarum (Sugarcane).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 OC NCBI_TaxID=4547;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Sugiharto B., Sakakibara H., Sugiyama T.;
 RA "Differential Expression of Two Genes for Sucrose-Phosphate Synthase
 RT in Sugarcane: Molecular Cloning of the cDNAs and Comparative Analysis
 RT of Gene Expression";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB001337; BAAL9241.1; -;
 DR GO; GO:0046524; F: sucrose-phosphate synthase activity; IEA.
 DR GO; GO:0016757; P: biosynthesis; IEA.
 DR GO; GO:0003058; P: biosynthesis; IEA.
 DR InterPro; IPR001296; Glyco_trans_1.
 DR Pfam; PF00534; Glycosyltransferase; 1.
 KW Glycosyltransferase; Transferase.
 FT NON TER. 1
 SQ SEQUENCE 1047 AA; 116379 MW; D0EDB34961E1D83D CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 1047;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVVPXGMDYS 11
 Db 414 VVPPGMDFS 422

RESULT 11

```
Q9SN30
ID Q9SN30 PRELIMINARY; PRT; 1083 AA.
AC Q9SN30;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
EN F28M11.40 OR AT4G10120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049487; CAB39784.1; -.
DR EMBL; AL161516; CAB78135.1; -.
DR PIR; T04062; T04062.
DR GO; GO:0046524; F:sucrose-phosphate synthase activity; IEA.
DR GO; GO:0016757; P:transferrase activity, transferring glycosyl...; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR Glycoyltransferase; Transferase.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1083 AA; 122688 MW; EFC126FCA2137BB1 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

2y 3 VVPXGMDYS 11
|:|:|:|:|
3b 483 VPPGMDFS 491

RESULT 12
Q43010 PRELIMINARY; PRT; 1084 AA.
AC Q43010;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sucrose phosphate synthase.
EN SP51.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
CX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shimada H.,
RA Fujimura T.;
RT "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS)
RT gene that is specifically expressed in the source organ."
RL Plant Sci. 112:207-217(1995).
DR EMBL; D45890; BAA08304.1; -.
DR PIR; T04103; T04103.
DR Gramene; Q43010; -.
DR GO; GO:0004459; F:L-lactate dehydrogenase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.

DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1100 AA; 121170 MW; 084F4604BA389CAD CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1100;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|
Db 469 VPPGMDFS 477

RESULT 14
Q9GQ04 PRELIMINARY; PRT; 219 AA.
AC Q9GQ04;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Na+/K+/2Cl-cotransporter (Fragment).
OS Eriocheir sinensis (Chinese mitten crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Grapsoidae; Varunidae; Eriocheir.
CX NCBI_TaxID=95602;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gill;
```

RA Wehrauch D., Towle D.W.;
 RT "Na+/H+-exchanger and Na+/K+/2Cl- -cotransporter are expressed in
 RL gills of the euryhaline Chinese crab *Eriocheir sinensis*.";

RL Comp. Biochem. Physiol. 126:SI58-SI58(2000).

DR ENBL; AF301160; AAG39938.1; -.

FT NON_TER 1

FT NON_TER 219 219

SQ SEQUENCE 219 AA; 24159 MW; 599442DA26F0D3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;

Best Local Similarity 75.0%; Pred. No. 49;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPXGMDYS 11

Db 107 VPOGLDYS 114

RESULT 15

Q8XPAB

ID Q8XPAB PRELIMINARY; PRT; 253 AA.

AC Q8XPAB;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein CPE0057.

GN CPE0057

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / Type A;

RX MEDLINE=21664373; PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,

RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic

RT flesh-eater";

RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

DR ENBL; AP003185; BAB79763.1; -.

DR GO; GO:0003824; Fcatalytic activity; IEA.

DR InterPro; IPR009036; Moeb.

DR InterPro; IPR00205; NAD_BS.

DR InterPro; IPR000594; ThiF domain.

DR Pfam; PF00899; ThiF; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 253 AA; 27714 MW; 814DF79D3E0D7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;

Best Local Similarity 50.0%; Pred. No. 58;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10

Db 108 EEIIPDDVDY 117

Search completed: June 3, 2004, 11:57:33

Job time : 29.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:11:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 BEVVPXGXHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	94.2	11	5	Abb80546 Hepatitis
2	49	94.2	11	5	Abb80554 Hepatitis
3	49	94.2	11	5	Abb80550 Hepatitis
4	49	94.2	11	5	Abb80555 Hepatitis
5	48	92.3	11	5	Abb80523 Hepatitis
6	48	92.3	11	5	Abb80558 Hepatitis
7	48	92.3	11	5	Abb80537 Hepatitis
8	48	92.3	11	5	Abb80560 Hepatitis
9	48	92.3	11	5	Abb80527 Hepatitis
10	48	92.3	11	5	Abb80541 Hepatitis
11	48	92.3	11	5	Abb80532 Hepatitis
12	48	92.3	11	5	Abb80531 Hepatitis
13	40	76.9	11	5	Abb80548 Hepatitis
14	40	76.9	11	5	Abb80549 Hepatitis
15	40	76.9	11	5	Abb80547 Hepatitis
16	40	76.9	11	5	Abb80544 Hepatitis
17	40	76.9	11	5	Abb80556 Hepatitis
18	40	76.9	11	5	Abb80557 Hepatitis
19	40	76.9	11	5	Abb80551 Hepatitis
20	40	76.9	11	5	Abb80553 Hepatitis
21	40	76.9	11	5	Abb80552 Hepatitis
22	40	76.9	11	5	Abb80545 Hepatitis
23	39	75.0	11	5	Abb80525 Hepatitis
24	39	75.0	11	5	Abb80534 Hepatitis
25	39	75.0	11	5	Abb80561 Hepatitis

26	39	75.0	11	5	Abb80521 Hepatitis
27	39	75.0	11	5	Abb80522 Hepatitis
28	39	75.0	11	5	Abb80536 Hepatitis
29	39	75.0	11	5	Abb80566 Hepatitis
30	39	75.0	11	5	Abb80542 Hepatitis
31	39	75.0	11	5	Abb80543 Hepatitis
32	39	75.0	11	5	Abb80563 Hepatitis
33	39	75.0	11	5	Abb80565 Hepatitis
34	39	75.0	11	5	Abb80524 Hepatitis
35	39	75.0	11	5	Abb80533 Hepatitis
36	39	75.0	11	5	Abb80529 Hepatitis
37	39	75.0	11	5	Abb80535 Hepatitis
38	39	75.0	11	5	Abb80567 Hepatitis
39	39	75.0	11	5	Abb80528 Hepatitis
40	39	75.0	11	5	Abb80538 Hepatitis
41	39	75.0	11	5	Abb80540 Hepatitis
42	39	75.0	11	5	Abb80562 Hepatitis
43	39	75.0	11	5	Abb80559 Hepatitis
44	39	75.0	11	5	Abb80526 Hepatitis
45	39	75.0	11	5	Abb80530 Hepatitis

ALIGNMENTS

RESULT 1

ABB80546
ID ABB80546 standard; peptide; 11 AA.

XX ABB80546;

AC AC

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

XX Synthetic.

XX Key

FH Key Location/Qualifiers

FT Modified-site 1

FT Modified-site 6 /note= "N-terminal acetyl"

FT Modified-site 6

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having Hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of

ABB80555
 ID ABB80555 standard; peptide; 11 AA.
 AC ABB80555;
 XT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
 CW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CX virucide.
 DS Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Misc-difference 8
 FT Modified-site 1 /note= "D-form residue"
 FT Modified-site 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 CX WO200208251-A2.
 PN 31-JAN-2002.
 PD 19-JUL-2001; 2001WO-US023169.
 PF 21-JUL-2000; 2000US-0220101P.
 CX (CORV-) CORVAS INT INC.
 CX Lim-Wilby M, Levy OE, Brunck TK;
 PX WPI; 2002-361643/39.
 CX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
 PX Claim 17; Page 65; 69pp; English.
 CX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
 SX Sequence 11 AA;
 Query Match 94.2%; Score 49; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0045;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGXHYS 11
 RESULT 5
 ABB80523
 ID ABB80523 standard; peptide; 11 AA.
 AC ABB80523;
 CX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CX virucide.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Misc-difference 9
 FT Modified-site 11 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 CX WO200208251-A2.
 PN 31-JAN-2002.
 PD 19-JUL-2001; 2001WO-US023169.
 PF 21-JUL-2000; 2000US-0220101P.
 CX (CORV-) CORVAS INT INC.
 CX Lim-Wilby M, Levy OE, Brunck TK;
 PX WPI; 2002-361643/39.
 CX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
 PX Claim 17; Page 64; 69pp; English.
 CX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
 SX Sequence 11 AA;
 Query Match 92.3%; Score 48; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0072;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHYS 11
 DB 1 EEVVPXGXHYS 11
 RESULT 6
 ABB80558
 ID ABB80558 standard; peptide; 11 AA.
 AC ABB80558;
 CX 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
 CW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 CX virucide.
 OS Synthetic.

PH	Key	Location/Qualifiers	
FT	Modified-site	1	
FT		/note= "N-terminal acetyl"	
FT	Modified-site	6	
FT		/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"	
FT	Modified-site	8	
FT		/note= "Oxymethionine"	
FT	Modified-site	11	
FT		/note= "C-terminal amide"	
XX			
XX	WO200208251-A2.		
XX			
XX	31-JAN-2002.		
XX			
XX	19-JUL-2001; 2001WO-US023169.		
XX			
XX	21-JUL-2000; 2000US-0220101P.		
PR			
XX	(CORV-) CORVAS INT INC.		
XX			
XX	Lim-Wilby M, Levy OE, Brunck TK;		
PI			
XX	WPI; 2002-361643/39.		
XX			
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.		
PT			
XX			
XX	Claim 17; Page 65; 69pp; English.		
XX			
XX	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus		
XX			
SQ	Sequence 11 AA:		
	Query Match	92.3%;	Score 48; DB 5; Length 11;
	Best Local Similarity	90.9%;	Pred. No; 0.0072;
	Matches 10; Conservative	0; Mismatches	1; Indels 0; Gaps 0
OY	1 EEVVPXGXHYS 11		
	1 EEVVPXGMHYS 11		
Db			
RESULT 7			
ABB80537			
ID	ABB80537 standard; peptide; 11 AA.		
XX			
XX	ABB80537;		
XX			
DT	08-OCT-2002 (first entry)		
XX			
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.		
XX			
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.		
KW			
XX	Synthetic.		
OS			
XX			
XX			
PH	Key	Location/Qualifiers	
FT	Modified-site	1	
FT		/note= "N-terminal acetyl"	
FT	Modified-site	6	
FT		/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"	
FT	Misc-difference	9	
FT		/note= "D-form residue"	
FT			

XX 31-JAN-2002.
PD
XX
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0072;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
2Y 1 EEVVPXGXHYS 11
DB ||||| |||||
1 EEVVPXGMHYS 11
RESULT 9
ID ABB80527
AC ABB80527 standard; peptide; 11 AA.
XX
XX ABB80527;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0072;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB ||||| |||||
1 EEVVPXGMHYS 11
RESULT 10
ID ABB80541
XX
XX ABB80541 standard; peptide; 11 AA.
XX
XX ABB80541;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.

XX PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.3%; Score 48; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0072;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11

DB 1 EEVVPXGXHYS 11

RESULT 11

ABB80532

ID ABB80532 standard; peptide; 11 AA.

XX AC ABB80532;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX KW virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Misc-difference 9 residue 7"

FT Modified-site 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX FN WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX PS Novel peptide compound having hepatitis C virus protease inhibitory

FT activity useful for treating disorders associated with hepatitis C virus

FT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.3%; Score 48; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0072;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11

DB 1 EEVVPXGXHYS 11

RESULT 12

ABB80531

ID ABB80531 standard; peptide; 11 AA.

XX AC ABB80531;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX KW virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Misc-difference 9 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

XX FN WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX PS Novel peptide compound having hepatitis C virus protease inhibitory

FT activity useful for treating disorders associated with hepatitis C virus

FT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.3%; Score 48; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0072;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11

Db 1 EVVFXGGHYS 11
|||||

RESULT 13

ABB80548
ID ABB80548 standard; peptide; 11 AA.

XX AC ABB80548;
XX

DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.
XX 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX CC Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 76.9%; Score 40; DB 5; Length 11;

Best Local Similarity 81.8%; Pred.No. 0.29;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVVFXGXHYS 11
|||||

DB 1 EVVFXGTDYS 11
|||||

RESULT 14

ABB80549
ID ABB80549 standard; peptide; 11 AA.

XX AC ABB80549;
XX

DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

AC ABB80549;

DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #29.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.
XX 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX CC Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 76.9%; Score 40; DB 5; Length 11;

Best Local Similarity 81.8%; Pred.No. 0.29;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVVFXGXHYS 11
|||||

DB 1 EVVFXGSSYS 11
|||||

RESULT 15

ABB80547

ID ABB80547 standard; peptide; 11 AA.

XX AC ABB80547;
XX

DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

```

XX OS Synthetic.
XX OS
XX PH Key Location/Qualifiers
XX FT Modified-site 1 /note= "N-terminal acetyl"
XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX FT Modified-site 11 residue 7"
XX FT Modified-site 11 /note= "C-terminal amide"
XX FT
XX PN WO200208251-A2.
XX XX
XX PD 31-JAN-2002.
XX XX
XX PF 19-JUL-2001; 2001WO-US023169.
XX XX
XX PR 21-JUL-2000; 2000US-0220101P.
XX XX
XX PA (CORV-) CORVAS INT INC.
XX XX
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX XX
XX DR WPI; 2002-361643/39.
XX XX
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX FT activity useful for treating disorders associated with hepatitis C virus
XX FT protease.
XX XX
XX PS Claim 17; Page 65; 69pp; English.
XX XX
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;

Query Match 76.9%; Score 40; DB 5; Length 11;
Best Local Similarity 81.8%; Pred. NO. 0.29;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
Db 1 EEVVPXGTDYS 11

Search completed: June 3, 2004, 11:48:24
Job time : 45.9333 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EEVVPXGXHY 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB pep:*

2: /cgn2_6/ptodata/2/iaa/5B COMB pep:*

3: /cgn2_6/ptodata/2/iaa/6A COMB pep:*

4: /cgn2_6/ptodata/2/iaa/6B COMB pep:*

5: /cgn2_6/ptodata/2/iaa/6C COMB pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	856	4 US-09-252-991A-21444	Sequence 21444, A
2	36	69.2	1037	4 US-09-134-001C-4794	Sequence 4794, A
3	35	67.3	277	4 US-09-252-991A-26615	Sequence 26615, A
4	34	65.4	126	2 US-08-879-995A-3	Sequence 3, Appli
5	34	65.4	126	3 US-09-215-096-3	Sequence 3, Appli
6	33	63.5	63	4 US-09-107-532A-6771	Sequence 6771, Ap
7	33	63.5	447	3 US-08-961-083-182	Sequence 182, App
8	33	63.5	447	4 US-09-536-784-182	Sequence 182, App
9	33	63.5	484	4 US-09-468-656A-6	Sequence 6, Appli
10	33	63.5	600	2 US-08-821-119-19	Sequence 19, Appli
11	33	63.5	600	2 US-08-821-118-2	Sequence 2, Appli
12	33	63.5	763	3 US-08-961-083-66	Sequence 66, Appli
13	33	63.5	763	4 US-09-536-784-66	Sequence 66, Appli
14	33	63.5	796	3 US-08-961-083-56	Sequence 56, Appli
15	33	63.5	796	4 US-09-536-784-56	Sequence 56, Appli
16	33	63.5	819	4 US-09-468-656A-8	Sequence 8, Appli
17	33	63.5	819	4 US-09-468-656A-10	Sequence 10, Appli
18	33	63.5	838	4 US-09-468-656A-4	Sequence 4, Appli
19	32	61.5	253	4 US-09-328-352-6440	Sequence 6440, Ap
20	32	61.5	323	4 US-09-543-681A-7304	Sequence 7304, Ap
21	32	61.5	738	1 US-08-530-010-3	Sequence 3, Appli
22	32	61.5	738	1 US-08-530-010-5	Sequence 5, Appli
23	32	61.5	738	1 US-08-530-010-7	Sequence 7, Appli
24	32	61.5	738	1 US-08-530-010-9	Sequence 9, Appli
25	32	61.5	738	1 US-08-530-010-11	Sequence 11, Appli
26	32	61.5	738	2 US-08-484-101B-3	Sequence 3, Appli
27	32	61.5	738	2 US-08-484-101B-5	Sequence 5, Appli

28 32 61.5 738 2 US-08-484-101B-7 Sequence 7, Appli
29 32 61.5 738 2 US-08-484-101B-9 Sequence 9, Appli
30 32 61.5 738 2 US-08-484-101B-11 Sequence 11, Appli
31 32 61.5 738 3 US-08-714-524D-3 Sequence 3, Appli
32 32 61.5 738 3 US-08-714-524D-5 Sequence 5, Appli
33 32 61.5 738 3 US-08-714-524D-7 Sequence 7, Appli
34 32 61.5 738 3 US-08-714-524D-9 Sequence 9, Appli
35 32 61.5 738 3 US-08-714-524D-11 Sequence 11, Appli
36 32 61.5 1247 4 US-09-252-991A-32960 Sequence 32960, A
37 31 59.6 70 4 US-09-134-001C-3950 Sequence 3950, Ap
38 31 59.6 237 4 US-09-134-001C-4185 Sequence 4185, Ap
39 31 59.6 252 4 US-09-328-352-7284 Sequence 7284, Ap
40 31 59.6 388 4 US-08-178-257-15 Sequence 15, Appli
41 31 59.6 390 4 US-09-489-039A-8065 Sequence 8065, Ap
42 31 59.6 502 4 US-09-342-647-4 Sequence 4, Appli
43 31 59.6 622 2 US-08-459-146-2 Sequence 2, Appli
44 31 59.6 622 2 US-08-459-065-2 Sequence 2, Appli
45 31 59.6 630 4 US-09-342-647-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-21444

; Sequence 21444, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,789

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21444

; LENGTH: 856

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21444

Query Match 73.1% Score 38; DB 4; Length 856;

Best Local Similarity 70.0%; Pred No. 18;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHY 10

Db 64 EAVVPGGHHY 73

RESULT 2

US-09-134-001C-4794

; Sequence 4794, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4794

; LENGTH: 1037

; TYPE: PRT

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.2%; Score 36; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHS 11
Db 199 KEVSNGLHYS 209

RESULT 3

US-09-252-991A-26615
Sequence 26615, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 26615

LENGTH: 277

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26615

Query Match

Best Local Similarity 67.3%; Score 35; DB 4; Length 277;

Mismatches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGXHS 11
Db 48 BEVPGGHTS 58

RESULT 4

US-08-879-995A-3

Sequence 3, Application US/08879995A

Patent No. 5985606

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Kaser, Matthew R.

TITLE OF INVENTION: HUMAN PREPROTACHYKININ B

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/879,995A

FILING DATE: Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 163590

US-08-879-995A-3

Query Match

Best Local Similarity 65.4%; Score 34; DB 2; Length 126;

Mismatches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXH 9
Db 28 EQVPGGGH 36

RESULT 5

US-09-215-096-3

Sequence 3, Application US/09215096

Patent No. 6008194

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Kaser, Matthew R.

TITLE OF INVENTION: HUMAN PREPROTACHYKININ B

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/215,096

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/879,995

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0326 US

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 163590

JG-09-215-096-3

Query Match 65.4%; Score 34; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DY 1 BEVVPXGXH 9
DB 28 EQVFGGHH 36

RESULT 6

US-09-107-532A-6771
; Sequence 6771, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6771:

SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...63
SEQUENCE DESCRIPTION: SEQ ID NO: 6771:
US-09-107-532A-6771

Query Match 63.5%; Score 33; DB 4; Length 63;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DY 1 BEVVPXGXHY 11
DB 5 EEMAFGDHYN 15

RESULT 7

US-08-961-083-182

; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 182:

SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 63.5%; Score 33; DB 3; Length 447;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DY 3 VVPXGXHY 10
DB 178 IVPHGGHY 185

RESULT 8

US-09-536-784-182
; Sequence 182, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match 63.5%; Score 33; DB 4; Length 447;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
Db 178 IVPHGGHY 185

RESULT 9
US-09-468-656A-6
Sequence 6, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
TITLE OF INVENTION: Motifs
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6

Query Match 63.5%; Score 33; DB 4; Length 484;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
Db 202 IVPHGGHY 209

RESULT 10
US-08-821-119-19
Sequence 19, Application US/08821119
Patent No. 5821104
GENERAL INFORMATION:
APPLICANT: Holm, Kaj Andre
APPLICANT: Rasmussen, Græthe
APPLICANT: Halkier, Torben
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: Tripeptidyl Aminopeptidase
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSER: NO. 58211040 No. 5821104disk of No. 5821104th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,119
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-119-19

Query Match 63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGXHY 11
Db 31 VPKGXHY 38

RESULT 11
US-08-821-118-2
Sequence 2, Application US/08821118
Patent No. 5989889
GENERAL INFORMATION:
APPLICANT: Rev, Michael
APPLICANT: Golightly, Elizabeth
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: NO. 59898890 No. 5989889disk of No. 5989889th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,118
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4107.400-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
JS-08-821-118-2

Query Match 63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2y 4 VPKGXHY 11
|||
31 VPKGWHY 38

RESULT 12
JS-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 63.5%; Score 33; DB 3; Length 763;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

2y 3 VVPXGXHY 10
:|:|:|

Db 184 IVPXGDHY 191

RESULT 13

US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

Query Match 63.5%; Score 33; DB 4; Length 763;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10
:|:|:|

Db 184 IVPXGDHY 191

RESULT 14

US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 63.5%; Score 33; DB 3; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 3 VVPXGXHY 10
Db 185 IVPXGDHY 192

RESULT 15

US-09-536-784-56
Sequence 56, Application US/09536784
Patent No. 6573082

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784

FILING DATE: 30-Oct-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56

Query Match 63.5%; Score 33; DB 4; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 3 VVPXGXHY 10
Db 185 IVPXGDHY 192

Search completed: June 3, 2004, 12:03:08
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

WM protein - protein search, using sw model

run on: June 3, 2004, 11:57:42 ; Search time 33 7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EEVVPXGXHYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	94.2	11	12	US-09-909-164-30
2	49	94.2	11	12	US-09-909-164-34
3	49	94.2	11	12	US-09-909-164-38
4	49	94.2	11	12	US-09-909-164-39
5	48	92.3	11	12	US-09-909-164-7
6	48	92.3	11	12	US-09-909-164-11
7	48	92.3	11	12	US-09-909-164-15
8	48	92.3	11	12	US-09-909-164-16
9	48	92.3	11	12	US-09-909-164-21
10	48	92.3	11	12	US-09-909-164-25
11	48	92.3	11	12	US-09-909-164-42
12	48	92.3	11	12	US-09-909-164-44
13	40	76.9	11	12	US-09-909-164-28
14	40	76.9	11	12	US-09-909-164-29
15	40	76.9	11	12	US-09-909-164-31

Sequence 32, Appl
Sequence 32, Appl
Sequence 35, Appl
Sequence 36, Appl
Sequence 37, Appl
Sequence 40, Appl
Sequence 41, Appl
Sequence 5, Appl
Sequence 6, Appl
Sequence 8, Appl
Sequence 9, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 14, Appl
Sequence 17, Appl
Sequence 18, Appl
Sequence 19, Appl
Sequence 20, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 24, Appl
Sequence 26, Appl
Sequence 27, Appl
Sequence 43, Appl
Sequence 45, Appl
Sequence 46, Appl
Sequence 47, Appl
Sequence 48, Appl
Sequence 49, Appl

ALIGNMENTS

RESULT 1

US-09-909-164-30

; Sequence 30, Application US/09909164

; Publication No. US20020086702A1

; GENERAL INFORMATION:

; APPLICANT: Corvas International, Inc.

; APPLICANT: Lim-Wilby, Marguerita

; APPLICANT: Levy, Odile E

; APPLICANT: Brunk, Terence K

; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C

; FILE REFERENCE: IN01192-US

; CURRENT APPLICATION NUMBER: US/09/909,164

; CURRENT FILING DATE: 2003-03-25

; PRIOR APPLICATION NUMBER: 60/220,101

; PRIOR FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 62

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 30

; LENGTH: 11

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: 11-mer synthesized according to example 1

; FEATURE:

; NAME/KEY: MOD_RES

; LOCATION: (1)..(1)

; OTHER INFORMATION: ACETYLATION

; FEATURE:

; NAME/KEY: MOD_RES

; LOCATION: (11)..(11)

; OTHER INFORMATION: AMIDATION

; FEATURE:

; NAME/KEY: MISC_FEATURE

; LOCATION: (6)..(6)

; OTHER INFORMATION: norvaline-(CO)

US-09-909-164-30

Query Match

94.2%; Score 49; DB 12; Length 11;

```
Best Local Similarity 90.3%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHYS 11
Db 1 EVVVPXGTHYS 11

RESULT 2
US-09-909-164-34
; Sequence 34, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION:
; US-09-909-164-34

Query Match 94.2%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHYS 11
Db 1 EVVVPXGSHYS 11

RESULT 3
US-09-909-164-38
; Sequence 38, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
```

```
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; US-09-909-164-38

Query Match 94.2%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHYS 11
Db 1 EVVVPXGSHYS 11

RESULT 4
US-09-909-164-39
; Sequence 39, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
```


; OTHER INFORMATION: D-amino acids
US-09-909-164-39

Query Match 94.2%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
| | | | | | | | | |
DB 1 EEVVPXGSHYS 11

RESULT 5

US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLYATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-11

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
| | | | | | | | | |
DB 1 EEVVPXGSHYS 11

US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLYATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-7

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
| | | | | | | | | |
DB 1 EEVVPXGSHYS 11

RESULT 6

US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US

; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLYATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-11

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
| | | | | | | | | |
DB 1 EEVVPXGSHYS 11

RESULT 7

US-09-909-164-15
; Sequence 15, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLYATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

QY 1 EEVVPXGXHYS 11
| | | | | | | | | |
DB 1 EEVVPXGSHYS 11

US-09-909-164-15
; Sequence 15, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLYATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)

```
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-15
Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
   |||||
Db 1 EEVVPXGXHYS 11

RESULT 8
US-09-909-164-16
; Sequence 16, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD_RES
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: PatentIn version 3.1
US-09-909-164-16
Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
   |||||
Db 1 EEVVPXGXHYS 11

RESULT 9
US-09-909-164-21
; Sequence 21, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
```

```
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: PatentIn version 3.1
US-09-909-164-21
Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
   |||||
Db 1 EEVVPXGXHYS 11

RESULT 10
US-09-909-164-25
; Sequence 25, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
```

```

; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
JS-09-909-164-25
Query Match      92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.0%; Pred.No. 0.006;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY 1 EEVVPXGXHYS 11
    |||||
Db 1 EEVVPXGXHYS 11

RESULT 11
US-09-909-164-42
; Sequence 42, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met(O)
US-09-909-164-44

Query Match      92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.006;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
    |||||
Db 1 EEVVPXGXHYS 11

RESULT 13
US-09-909-164-28
; Sequence 28, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met(O)
US-09-909-164-42

Query Match      92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.006;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
    |||||
Db 1 EEVVPXGXHYS 11

RESULT 12
US-09-909-164-44
; Sequence 44, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
```

OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
US-09-909-164-28

Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
|||||
Db 1 BEVVPXGTSYS 11

RESULT 14
US-09-909-164-29
Sequence 29, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 29
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
US-09-909-164-29

Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
|||||
Db 1 BEVVPXGTSYS 11

RESULT 15
US-09-909-164-31
Sequence 31, Application US/09909164

Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunk, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 31
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
US-09-909-164-31

Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
|||||
Db 1 BEVVPXGTSYS 11

Search completed: June 3, 2004, 12:57:16
Job time : 33.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EVVVPXGXHY 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35	67.3	233	2 T02590	DNA binding protei
2	35	67.3	308	2 A72207	ftSH proteinase ac
3	35	67.3	360	2 E63086	cell division prot
4	35	67.3	1028	2 AF3286	Arp-dependent DNA
5	35	67.3	1396	2 S36851	L-shaped tail fibe
6	34	65.4	102	2 A42452	V1 protein - tobac
7	34	65.4	126	2 A25905	tachykinin B precu
8	34	65.4	279	2 C75538	hypothetical prote
9	34	65.4	495	2 T28717	hypothetical prote
10	34	65.4	534	2 A69284	coenzyme F420-quin
11	34	65.4	822	2 T46758	hypothetical 92.4K
12	34	65.4	1057	2 F89892	carbamoyl-phosphat
13	33	63.5	46	2 B97985	hypothetical prote
14	33	63.5	124	1 VKLJ51	trans-regulatory s
15	33	63.5	156	2 D82618	conserved hypothet
16	33	63.5	252	2 AE2001	hypothetical prote
17	33	63.5	412	2 A48702	2-methyl-branched
18	33	63.5	460	2 S69046	hypothetical prote
19	33	63.5	510	2 G86430	TS18.1 protein - A
20	33	63.5	743	2 S38143	hypothetical prote
21	33	63.5	802	2 C95136	conserved domain p
22	33	63.5	819	2 B95136	conserved domain p
23	33	63.5	828	2 E98004	hypothetical prote
24	33	63.5	839	2 G95115	conserved hypothet
25	33	63.5	853	2 C97985	hypothetical prote
26	33	63.5	855	2 D98004	histidine Motif-Co
27	33	63.5	1039	2 H95115	conserved hypothet
28	33	63.5	1039	2 D97985	hypothetical prote
29	32	61.5	162	2 A70939	probable monoxigen

30 32 61.5 225 2 S57810
31 32 61.5 286 2 C86169
32 32 61.5 311 2 H69194
33 32 61.5 322 2 AF3211
34 32 61.5 364 2 S47540
35 32 61.5 369 2 T17267
36 32 61.5 374 2 G69119
37 32 61.5 425 2 T24111
38 32 61.5 454 2 G96568
39 32 61.5 519 2 G84598
40 32 61.5 536 2 AI3544
41 32 61.5 627 2 A69863
42 32 61.5 696 2 A91247
43 32 61.5 701 2 S61239
44 32 61.5 716 1 JC5061
45 32 61.5 738 2 A48246

ALIGNMENTS

RESULT 1

T02590

DNA binding protein EREBP-2 - common tobacco

C/Species: Nicotiana tabacum (common tobacco)

C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000

C/Accession: T02590

R/Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A/Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi

A/Reference number: Z14671; MUID:95276459; PMID:7756928

A/Accession: T02590

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-233 <OHM>

A/Cross-references: EMBL:D38126; NID:G790362; PIDN:BAA07324.1; PID:G1208498

A/Experimental source: strain BY4; tissue-type leaf

Query Match 67.3%; Score 35; DB 2; Length 233;

Best Local Similarity 60.0%; Pred. No. 9.8;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHY 10

Db 90 QAVVPKGRHY 99

RESULT 2

A72207

ftSH proteinase activity modulator HfLK - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C/Accession: A72207

R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;

C.M.

Nature 399, 323-329, 1999

A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: A72207

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-308 <ARN>

A/Cross-references: GB:AE001819; GB:AE000512; NID:G4982396; PIDN:AD36885.1; PID:G498240

A/Experimental source: strain MSB8

C/Genetics:

A/Gene: TM1822

C/Superfamily: erythrocyte band 7 integral membrane protein

Query Match 67.3%; Score 35; DB 2; Length 308;

Best Local Similarity 75.0%; Pred. No. 13;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 3 VVPKXGHY 10
Db 41 VVPSGHHY 48

RESULT 3
E69086
cell division protein - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: E69086
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
  Qiu, D.; Spadafora, R.; Vicair, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
  ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:93711463
A:Accession: E69086
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360 <MTH>
A:Cross-references: GB:AE000923; GB:AE000666; NID:g2622766; PIDN:AAB86115.1; PID:g262277
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1642
C:Superfamily: cell division protein MJ0174

Query Match 67.3%; Score 35; DB 2; Length 360;
Best Local Similarity 45.5%; Pred. No. 16;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPKXGHYS 11
Db 98 EDLVPVGSHT 108

RESULT 4
AF3286
ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3286
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
  Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1028 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51457.1; PID:g17982167; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0275
A:Map position: 1

Query Match 67.3%; Score 35; DB 2; Length 1028;
Best Local Similarity 54.5%; Pred. No. 49;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPKXGHYS 11
Db 76 EKIVPPGARYS 86

RESULT 5
S3681
L-shaped tail fiber protein - phage T5
N:Alternate names: ltf protein
C:Species: phage T5
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C:Accession: S65934; S01984; S3681

R:Kallman, A.V.; Kulshin, V.E.; Shlyapnikov, M.G.; Ksenzenko, V.N.; Kryukov, V.M.
FEBS Lett. 366, 46-48, 1995
A:Title: The nucleotide sequence of the bacteriophage T5 ltf gene.
A:Reference number: S65934; MUID:95309401; PMID:7789514
A:Accession: S65934
A:Molecule type: DNA
A:Residues: 1-1396 <KAL>
A:Cross-references: EMBL:X69460; NID:g15415; PIDN:CAA49220.1; PID:g15416
R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
Nucleic Acids Res. 16, 6230, 1988
A:Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early an
A:Reference number: S01982; MUID:88289370; PMID:3267228
A:Accession: S01984
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 934-985, 'A', 987-1396 <KA2>
A:Cross-references: EMBL:X07559
C:Genetics:
A:Gene: ltf
C:Keywords: late protein; tail fiber

Query Match 67.3%; Score 35; DB 2; Length 1396;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVPKXGHYS 11
Db 1360 KTIPAGDHYS 1369

RESULT 6
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Hailey, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yell
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:g35283; PIDN:AAA47947.1; PID:g335284

Query Match 65.4%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 6.5;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPKXGHYS 11
Db 7 QVVPSGINYS 16

RESULT 7
A25905
tachykinin B precursor - bovine
N:Alternate names: neuromedin K
C:Species: Bos primigenius taurus (cattle)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25905
R:Kotani, H.; Hoehmaru, M.; Nawa, H.; Nakanishi, S.
Proc. Natl. Acad. Sci. U.S.A. 83, 7074-7078, 1986
A:Title: Structure and gene organization of bovine neuromedin K precursor.
A:Reference number: A25905; MUID:86313713; PMID:3462746
A:Accession: A25905
A:Molecule type: DNA
A:Residues: 1-126 <KOT>
A:Cross-references: GB:M14351; NID:g163587; PIDN:AAA30723.1; PID:g163590
C:Superfamily: neurokinin B precursor
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-126/Product: protachykinin B #status predicted <MAT>

```

```

Query Match      65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 8.2;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 1 EVVVPXGXH 9
Db 28 EQVVPGGH 36

RESULT 8
C75538
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: C75538
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75538
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <WHI>
A:Cross-references: GB:AE001889; GB:AE00513; NID:G6457944; PIDN:AAF09867.1; PID:G645795
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0271
A:Map position: 1
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match      65.4%; Score 34; DB 2; Length 279;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 4 VPKXGXHS 11
Db 100 VPLGRHS 107

RESULT 9
T28717
hypothetical protein F10D2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T28717
R:Graves, T.; Wohlmann, P.; Gillam, B.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid F10D2.
A:Reference number: Z20515
A:Accession: T28717
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-495 <GRA>
A:Cross-references: EMBL:AF022972; PIDN:AAC48234.1; GSPDB:GN00023; CESP:F10D2.3
A:Experimental source: strain Bristol N2; clone F10D2
C:Genetics:
A:Gene: CESP:F10D2.3
A:Map position: 5
A:Introns: 37/2; 90/1; 113/1; 183/3; 356/3; 380/3; 428/2

Query Match      65.4%; Score 34; DB 2; Length 495;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXHY 10
Db 218 ENIVTGRKH 227

RESULT 10
A69284
coenzyme F420-quinone oxidoreductase (EC 1.6.5.-) 56K chain - Archaeoglobus fulgidus
N:Alternate names: sarcosine oxidase alpha chain soxA homolog
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000
C:Accession: A69284; S45665
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69284
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-534 <KLE>
A:Cross-references: GB:AE001086; GB:AE000782; NID:G2689409; PIDN:AAB90959.1; PID:G26503;
R:Kunow, J.; Linder, D.; Stetter, K.O.; Thauer, R.K.
Eur. J. Biochem. 223, 503-511, 1994
A>Title: F(420)H(2): quinone oxidoreductase from Archaeoglobus fulgidus. Characterizati
A:Reference number: S45665; MUID:94333340; PMID:8055920
A:Accession: S45665
A:Molecule type: protein
A:Residues: 2,'X',4,'X',6-7,'X',9-10,'XX',13-14 <KUN>
A>Note: the authors could not distinguish between glutamate and cysteine for residues 5-
C:Superfamily: glutamate synthase small chain
C:Keywords: oxidoreductase

Query Match      55.4%; Score 34; DB 2; Length 534;
Best Local Similarity 45.5%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXHS 11
Db 119 DKVFPGHSHY 129

RESULT 11
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heynemann, J.; Schnitzler, N.; Luet
Infest. Immun. 67, 871-878, 1999
A>Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachme
A:Reference number: Z24091; MUID:99115568; PMID:9916102
A:Accession: T46758
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPE>
A:Cross-references: EMBL:AF062533; NID:G4249622; PIDN:AAD13797.1; PID:G4249624
A:Experimental source: strain R268
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match      55.4%; Score 34; DB 2; Length 822;
Best Local Similarity 75.0%; Pred. No. 62;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
Db 350 VVPXGXHY 357

RESULT 12
F89892
carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89892
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

```

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A;Reference number: A89758; MUID:21311952; PMID:11418146
 A;Accession: F89822
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1057 <KUR>
 A;Cross-references: GB:BA000018; PID:g13701002; PIDN:BA842298.1; GSPDB:GN00149
 A;Experimental source: strain N315
 C;Genetics:
 A;Gene: pyrAB
 C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 65.4%; Score 34; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 82;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYS 11
 ||| |||
 Db 190 EIVSNGLHYS 199
 ||| |||

RESULT 13
 E97985
 hypothetical protein spr0909 [imported] - Streptococcus pneumoniae (strain R6)
 C;Species: Streptococcus pneumoniae
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C;Accession: E97985
 R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E
 r, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:11544234
 A;Accession: E97985
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-46 <KUR>
 A;Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:g15458516; GSPDB:GN00174
 C;Genetics:
 A;Gene: spr0909

Query Match 63.5%; Score 33; DB 2; Length 46;
 Best Local Similarity 62.5%; Pred. No. 4.5;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 ||| |||
 Db 21 IVPRGGHY 28
 ||| |||

RESULT 14
 VKLJSI
 trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz
 N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs protein
 C;Species: simian immunodeficiency virus SIVcpz
 A;Note: host Pan troglodytes (chimpanzee)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: S09988
 R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A;Reference number: S09983; MUID:90259077; PMID:2188136
 A;Accession: S09988
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-124 <HUE>
 A;Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:g763085
 C;Genetics:
 A;Gene: rev; trs; art
 A;Introns: 27/1

C;Superfamily: AIDS trans-regulatory splicing protein
 C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYS 11
 ||| |||
 Db 107 EIVPAGGNTS 116
 ||| |||

RESULT 15
 D82618
 conserved hypothetical protein XF1950 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C;Species: *Xylella fastidiosa*
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C;Accession: D82618
 R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A;Reference number: A82515; MUID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: D82618
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-156 <SIM>
 A;Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GN00
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.B.A.; Camargo, D.M.; Carret,
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasa
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
 A;Reference number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: XF1950

Query Match 63.5%; Score 33; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGXH 9
 ||| |||
 Db 119 EETLPQGVH 127
 ||| |||

Search completed: June 3, 2004, 12:00:01
 Job time : 9 secs

RESULT 2

CARB STAPP
ID CARB STAPP STANDARD; PRT; 1057 AA.
AC Q8CPJ4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain)
DE phosphate synthetase ammonia chain)
GN CARB OR SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J., Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z., Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RA "Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228).";
RT Mol. Microbiol. 49:1577-1593(2003).
RL -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).
CC EMBL; AE016746; AAC0476.1; --
DR HAMAP; MF 01210; -- 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS like.
DR InterPro; IPR000169; SHPOT.acsite.
DR Pfam; PF02089; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRfams; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_2; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
FT ATP-binding; Manganese; Complete proteome.
KW CARBOXYPHOSPHATION DOMAIN.
FT DOMAIN 1 401
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1057
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352
FT NP_BIND 284 284
FT METAL 284 298
FT METAL 298

FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;
Query Match 69.2%; Score 36; DB 1; Length 1057;
Best Local Similarity 63.6%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 BEVVPXGXHY 11
Db 189 KEVVSNGLHY 199
RESULT 3
CARB FUSNN STANDARD; PRT; 1058 AA.
ID CARB FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DS Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
DE phosphate synthetase ammonia chain).
GN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21863394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Barman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein M., Kyrides N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586";
RL J. Bacteriol. 184:2005-2018(2002).
RC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).
CC EMBL; AE010554; AAL94625.1; ALT_INIT.
DR HAMAP; MF 01210; -- 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS like.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRfams; TIGR01369; CPSaseII_lrg; 1.


```

RN SEQUENCE FROM N.A.
RP MEDLINE=96313713; PubMed=3462746;
RX Kotani H., Hoshimaru M., Nawa H., Nakanishi S.;
RA "Structure and gene organization of bovine neuromedin K precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7074-7078 (1986).
CC -!- FUNCTION: Tachykinins are active peptides which excite neurons,
CC evoke behavioral responses, are potent vasodilators and
CC secretagogues, and contract (directly or indirectly) many smooth
CC muscles
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the tachykinin family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14351; AAA30723.1; JOINED.
CC EMBL; M14347; AAA30723.1; JOINED.
CC EMBL; M14348; AAA30723.1; JOINED.
CC EMBL; M14349; AAA30723.1; JOINED.
CC EMBL; M14350; AAA30723.1; JOINED.
CC PIR; A25905; A25905.
CC InterPro; IPR003635; Neurokinin.
CC InterPro; IPR002040; Tachy Neurokinin.
CC Pfam; PF03823; Neurokinin B; 1.
CC ProDom; PD020370; Neurokinin; 1.
CC Tachykinin; Neurokinin; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 83 NEUROKININ B.
FT PEPTIDE 86 95
FT PROPEP 99 126
FT MOD RES 95 95
FT SEQUENCE 126 AA; 13871 MW; 4468F433498EC059 CRC64;
Query Match 55.4%; Score 34; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 BEVVPXGXH 9
Db 28 EQVFGGHH 36
RESULT 7
CARB STAAW STANDARD; PRT; 1057 AA.
AC Q99UR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR SAV1203 OR SA1046.
OS Staphylococcus aureus (strain Wd50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN SEQUENCE FROM N.A.
RP STRAIN=Md50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,

```

```

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240 (2001).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carB family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003361; BAB57365.1; -.
CC EMBL; AP003132; BAB42298.1; -.
CC PIR; F89892; F89892.
CC HSSP; P00968; ICS0.
CC HAMAP; MF_01210; -; 1.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS_like.
CC Pfam; PF02789; CPase_L_chain; 2.
CC Pfam; PF02786; CPase_L_D2; 2.
CC Pfam; PF02787; CPase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPASE.
CC TIGRfam; TIGR01389; CPaseII_lrg; 1.
CC PROSITE; PS00866; CPASE_1; 2.
CC PROSITE; PS00867; CPASE_2; 2.
CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;
Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVFXGXHYS 11
Db 190 EIVSNGLHYS 199
RESULT 8
CARB STAAW STANDARD; PRT; 1057 AA.
ID CARB STAAW
AC P58940;

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain)
GN CARB OR PYRAB OR MW1086.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwana N., Asano K., Naimi T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.;
RA "Genome and virulence determinants of high virulence community-acquired MRSA";
RT Lancet 359:1819-1827(2002).
RL
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AF004825; BAB94951.1; -
DR HAMAB; MF_01210; -; 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Chase_L.
DR InterPro; IPR005479; Chase_L_D2.
DR InterPro; IPR005480; Chase_L_D3.
DR InterPro; IPR005481; Chase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF02089; CFSase_L_chain; 2.
DR Pfam; PF02786; CFSase_L_D2; 2.
DR Pfam; PF02787; CFSase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CFSASE.
DR TIGRfams; TIGR01369; CFSaseII_lrg; 1.
DR PROSITE; PS00866; CFSASE_1; 2.
DR PROSITE; PS00867; CFSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat; ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401
FT CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546
FT OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929
FT CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057
FT ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT ATP (POTENTIAL).
FT NP_BIND 302 352
FT MANGANESE 1 (BY SIMILARITY).
FT METAL 284 284
FT MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 298 298
FT MANGANESE 2 (BY SIMILARITY).
FT METAL 300 300
FT MANGANESE 3 (BY SIMILARITY).
FT METAL 820 820
FT MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832
FT MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYS 11
Db 190 EIVSNGLHYS 199
RESULT 9
ID REV_SIVCZ STANDARD; PRT; 124 AA.
AC PF7280;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE REV protein (Anti-repression transactivator protein) (ART/TRS).
GN REV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90259077; PubMed=2188136;
RA Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
RA "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
RL Nature 345:356-359(1990).
CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
CC -!- PTM: Phosphoprotein whose state of phosphorylation is mediated by a specific serine kinase activity present in the nucleus.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X52154; CAA36405.1; -
DR FIR; S09988; VKLJSI.
DR HIV; X52154; REVSCP2.
DR InterPro; IPR000625; REV_protein.
DR Pfam; PF00424; REV; 1.
KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
SQ SEQUENCE 124 AA; 13701 MW; F5877D1BD65A7B2 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 124;
Best Local Similarity 60.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYS 11
Db 107 ETVPAGGNYS 116
RESULT 10
ID UMEI_YEAST STANDARD; PRT; 460 AA.
AC Q03010; P87330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Meiosis negative regulator UME1.
GN UME1 OR WTM3 OR YPL139C OR LPI7C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364A;
RA Mallory M.J., Strich R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=S288C / AB972;
RX  MEDLINE=97313271; PubMed=9169875;
RA  Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Anasorge W.,
RA  Araujo R., Aparicio A., Barrill B.G., Badcock K., Benes V.,
RA  Bostein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA  Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA  Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA  Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA  Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA  Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA  Komp C., Kurdi O., Lahekari D., Lew H., Lin A., Lin D., Louis E.J.,
RA  Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA  Mueller-Auer S., Namath A., Nenwich U., Oefner P., Pearson D.,
RA  Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA  Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA  Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA  Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA  Zhong W., Zollner A., Vo D.H., Hani J.,
RL  "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL  Nature 387:103-105(1997).
CC  -!- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC  REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC  -!- SIMILARITY: Contains 4 WD repeats.
CC  -!- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; U10280; AB40937.1; -
CC  DR  EMBL; U43703; AB68221.1; -
CC  DR  F01; S69046; S69046.
CC  DR  GenOnline; 144121; -
CC  DR  TRANSFAC; T04309; -
CC  DR  SGD; S0006060; UME1.
CC  DR  GO; GO:0005634; C:nucleus; IDA.
CC  DR  GO; GO:0005714; C:transcription co-repressor activity; IDA.
CC  DR  GO; GO:0040020; P:regulation of meiosis; IGI.
CC  DR  InterPro; IPR001680; WD40.
CC  DR  Pfam; PF00400; WD40; 3.
CC  DR  SMART; SM00320; WD40; 4.
CC  DR  PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
CC  DR  PROSITE; PS00082; WD_REPEATS_2; FALSE_NEG.
CC  DR  PROSITE; PS02994; WD_REPEATS_REGION; FALSE_NEG.
KW  Transcription regulation; Meiosis; Repeat; WD repeat.
FT  REPEAT 233 271
FT  REPEAT 276 316
FT  REPEAT 339 379
FT  REPEAT 411 451
FT  REPEAT 451 451
SQ  SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
DB 85 IVPGLGHI 92

RESULT 11
ID MDLB BUCBP STANDARD; PRT; 578 AA.
AC Q89A36;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Multidrug resistance-like ATP-binding protein mdlb.
GN MDLB OR BBP424.
OC Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva P.J.,
RA Tamaras J., Viqueira E., Latorre A., Valencia A., Moran F., Moya A.;
RA "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB014017; AA027134.1; -
CC DR  InterPro; IPR001140; ABC_TM_transpt.
CC DR  InterPro; IPR003439; ABC_transporter.
CC DR  Pfam; PF00664; ABC_membrane_1.
CC DR  Pfam; PF00005; ABC_tran; 1.
CC DR  PROSITE; PS00929; ABC_TMIF; 1.
CC DR  PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
CC DR  PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC DR  ATP-binding; Transport; Transmembrane; Complete proteome.
CC FT  TRANSMEM 26 46
CC FT  TRANSMEM 59 79
CC FT  TRANSMEM 143 163
CC FT  TRANSMEM 166 186
CC FT  TRANSMEM 196 216
CC FT  TRANSMEM 260 280
CC FT  DOMAIN 339 573
CC FT  NP_BIND 373 380
CC FT  ATP (POTENTIAL).
CC SQ  SEQUENCE 578 AA; 66827 MW; 3B84848CB196ADF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
DB 550 EIVESGTHN 559

RESULT 12
YK47 YEAST STANDARD; PRT; 743 AA.
AC P36148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCPI-MET1 intergenic region.
GN YKR067W.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA van Vliet-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: STRONG, TO YEAST YBL011W.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z28292; CAAB2146.1; -;
 CC PIR; S38143; S38143.
 CC Germonline; 140046; -;
 CC SGD; S0001775; GPR2.
 CC GO; GO:0005783; C:cytoplasmic reticulum; IDA.
 CC GO; GO:0043666; F:glycerol-3-phosphate O-acyltransferase acti. . .; IDA.
 CC GO; GO:008654; P:phospholipid biosynthesis; IDA.
 CC InterPro; IPR002123; Acyltransferase.
 CC Pfam; PF01553; Acyltransferase; 1.
 CC SMART; SM00563; PlsC; 1.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 31 55 POTENTIAL.
 CC FT TRANSMEM 69 85 POTENTIAL.
 CC FT TRANSMEM 502 524 POTENTIAL.
 CC FT TRANSMEM 539 555 POTENTIAL.
 CC SEQUENCE 743 AA; 83644 MW; 8489946E56B82F15 CRC64;
 CC
 CC Query Match 63.5%; Score 33; DB 1; Length 743;
 CC Best Local Similarity 75.0%; Pred. No. 43;
 CC Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 3 VVPXGXHY 10
 CC DB 294 VVPCGLHY 301
 CC
 CC RESULT 13
 CC MSRA_VIBPA
 CC ID MSRA_VIBPA STANDARD; PRT; 212 AA.
 CC AC Q87SW6;
 CC DT 10-OCT-2003 (Rel. 42, Created)
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Peptide methionine sulfoxide reductase msrA (EC 1.8.4.6) (Protein-
 CC methionine-S-oxide reductase) (Peptide Met (O) reductase).
 CC GN MSRA OR VP0306.
 CC OS Vibrio parahaemolyticus.
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC OC Vibrionaceae; Vibrio.
 CC OX NCBI_TaxID=670;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=BM2210633 / Serotype O3:K6;
 CC RX MEDLINE=22508454; PubMed=12620739;
 CC RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 CC RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 CC RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 CC RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 CC distinct from that of *V. cholerae*.";
 CC RL Lancet 361:743-749(2003).
 CC
 CC -!- FUNCTION: Has an important function as a repair enzyme for
 CC proteins that have been inactivated by oxidation. Catalyzes the
 CC reversible oxidation-reduction of methionine sulfoxide in proteins
 CC to methionine (By similarity).
 CC
 CC -!- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin =
 CC protein L-methionine S-oxide + reduced thioredoxin.
 CC
 CC -!- SIMILARITY: Belongs to the msrA Met sulfoxide reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AP005074; BACS6569.1; -;
 CC HAMAP; MF_01401; -; 1.
 CC InterPro; IPR002569; PMSR.
 CC Pfam; PF01625; PMSR; 1.
 CC KW Oxidoreductase; Complete proteome.
 CC FT ACT_SITE 51 51 BY SIMILARITY.
 CC SEQUENCE 212 AA; 23616 MW; 1967AFEC3C735EC6 CRC64;
 CC
 CC Query Match 61.5%; Score 32; DB 1; Length 212;
 CC Best Local Similarity 55.6%; Pred. No. 19;
 CC Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 2 EWPXGXHY 10
 CC DB 171 EIVPAGPY 179
 CC
 CC RESULT 14
 CC GAAB_METH
 CC ID GAAB_METH STANDARD; PRT; 308 AA.
 CC AC Q26806;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP
 CC synthetase).
 CC GN GUAAB OR MTH710.
 CC OS Methanobacterium thermoautotrophicum.
 CC OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 CC OC Methanobacteriaceae; Methanothermobacter.
 CC OX NCBI_TaxID=187420;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=Delta H;
 CC RX MEDLINE=98037514; PubMed=9371463;
 CC RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 CC RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
 CC RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 CC RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 CC RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 CC RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 CC RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 CC RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*
 CC deltaH: functional analysis and comparative genomics.";
 CC RL J. Bacteriol. 179:7135-7155(1997).
 CC
 CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
 CC
 CC -!- SUBUNIT: Heterodimer composed of a glutamine amidotransferase
 CC subunit (A) and a GMP synthase subunit (B) (Potential).
 CC
 CC -!- SIMILARITY: Belongs to the GMP synthase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000850; AAB85215.1; ALT_INIT.
 CC HAMAP; MF_00345; -; 1.
 CC HSSP; P04079; 1GPM.
 CC InterPro; IPR001674; GMP_synth_C.
 CC Pfam; PF00958; GMP_synth_C; 1.
 CC TIGRFAMs; TIGR00884; GnaA_Cterm; 1.
 CC KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 CC Complete proteome.
 CC FT DOMAIN 33 184
 CC FT NP_BIND 29 35 ATP-BINDING (BY SIMILARITY).
 CC SEQUENCE 308 AA; 34403 MW; F2DCF6ED202CAEC1 CRC64;
 CC SQ

Query Match 61.5%; Score 32; DB 1; Length 308;
Best Local Similarity 63.6%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EVVPXGXHYS 11
|||
DB 216 EEVVEGLHES 226

Db 189 EVIPDGS 196
||:|

Search completed: June 3, 2004, 11:49:54
Job time : 5.86667 secs

RESULT 15
ALFB_SHEEP
ID ALFB_SHEEP STANDARD; PRT; 363 AA.
AC PS2210;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).
GN ALDOB.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mesonephros;
RX MEDLINE=94368863; PubMed=8086469;
RA Gianquinto L., Pailhoux E.A., Bezard J., Servel N., Kirszenbaum M.,
RA Cotinot C.;
RT "Cloning and characterization of a full-length cDNA coding for ovine
aldolase B from fetal mesonephros";
RL Biochim. Biophys. Acta 1219:223-227 (1994).
CC -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
phosphate + D-glyceraldenide 3-phosphate.
CC -!- PATHWAY: Glycolysis; sixth step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- MISCELLANEOUS: In vertebrates, three forms of this ubiquitous
glycolytic enzyme are found, aldolase A in muscle, aldolase B in
liver and aldolase C in brain.
CC -!- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; 229372; CAA82563.1; -;
DR PIR; S47540; S47540.
DR HSP; P00893.1; DO.
DR InterPro; IPR000741; Aldolase_I.
DR Pfam; PF00274; Glycolytic_ery; 1.
DR ProDom; PD001128; Aldolase_I; 1.
DR PROSITE; PS00158; ALDOLASE_CLASS_I; 1.
KW Lyase; Schiff base; Glycolysis; Multigene family.
FT INIT MET 0
FT BINDING 55 55 BY SIMILARITY.
FT BINDING 146 146 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT BINDING 229 229 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT ACT_SITE 363 363 SCHIFF-BASE WITH DIHYDROXYACETONE-P.
FT ACT_SITE 363 363 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
SQ SEQUENCE 363 AA; 39500 MW; FC8B4566821E2BD CRC64;

Query Match 61.5%; Score 32; DB 1; Length 363;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXH 9

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EEVVPXGXHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organalle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	78.8	413	11 Q8K289	Q8K289 mus musculus
2	36	69.2	208	2 Q46486	Q46486 corynebacte
3	36	69.2	1057	16 Q8CPJ4	Q8CPJ4 staphylococ
4	35	67.3	233	10 Q40479	Q40479 nicotiana t
5	35	67.3	237	10 Q9LW50	Q9LW50 nicotiana s
6	35	67.3	285	16 Q98H06	Q98H06 rhizobium 1
7	35	67.3	308	16 Q9X2E2	Q9X2E2 thermotoga
8	35	67.3	317	9 Q38317	Q38317 lactobacill
9	35	67.3	360	17 Q27679	Q27679 methanobact
10	35	67.3	545	10 Q7XTG3	Q7XTG3 cryza sativ
11	35	67.3	678	12 Q9ELX6	Q9ELX6 cercopithe
12	35	67.3	855	5 Q8IME6	Q8IME6 drosophila
13	35	67.3	1028	16 Q8YJ11	Q8YJ11 bruceella me
14	35	67.3	1044	16 Q8DIH0	Q8DIH0 synchococ
15	34	65.4	262	16 Q7WN57	Q7WN57 bordetella
16	34	65.4	262	16 Q7W023	Q7W023 bordetella

17	34	65.4	279	16 Q9RXN9	Q9RXN9 deinococcus
18	34	65.4	322	17 Q9HLH8	Q9HLH8 thermoplasm
19	34	65.4	495	5 O16912	O16912 caenorhabdi
20	34	65.4	534	17 O29966	O29966 archaeoglob
21	34	65.4	794	16 Q877Y2	Q877Y2 streptococc
22	34	65.4	822	2 Q9ZHG7	Q9ZHG7 streptococc
23	34	65.4	822	16 Q8B4U1	Q8B4U1 streptococc
24	34	65.4	822	16 Q8DZ81	Q8DZ81 streptococc
25	34	65.4	823	16 Q8N282	Q8N282 streptococc
26	34	65.4	823	16 Q8K5Q1	Q8K5Q1 streptococc
27	34	65.4	825	16 Q99XV4	Q99XV4 streptococc
28	34	65.4	962	5 Q8IB11	Q8IB11 plasmodium
29	34	65.4	1442	17 Q96VH5	Q96VH5 sulfolobus
30	33	63.5	46	16 Q8CVU7	Q8CVU7 streptococc
31	33	63.5	78	6 Q9XST4	Q9XST4 canis famil
32	33	63.5	139	2 Q57489	Q57489 bacteroides
33	33	63.5	156	16 Q9PC35	Q9PC35 xyiella fas
34	33	63.5	156	16 Q87D36	Q87D36 xyiella fas
35	33	63.5	175	6 Q8I033	Q8I033 bos taurus
36	33	63.5	215	6 Q8I031	Q8I031 bos taurus
37	33	63.5	217	4 Q00404	Q00404 homo sapien
38	33	63.5	241	16 Q87XD7	Q87XD7 pseudomonas
39	33	63.5	252	16 Q8TWFL	Q8TWFL anabaena sp
40	33	63.5	281	6 Q867A5	Q867A5 tragus ja
41	33	63.5	297	6 Q8HXY9	Q8HXY9 bos taurus
42	33	63.5	297	16 Q7VP43	Q7VP43 haemophilus
43	33	63.5	299	4 Q9URE9	Q9URE9 homo sapien
44	33	63.5	364	5 Q8IGN5	Q8IGN5 drosophila
45	33	63.5	412	5 Q08523	Q08523 ascaris suu

ALIGNMENTS

RESULT 1

Q8K289 PRELIMINARY; PRT; 413 AA.

AC Q8K289;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to expressed sequence A1987856.
GN TADA3L OR 1110004B19RIK.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC032195; AAH32195.1; -

DR MGD; MGI:1915724; Tada3l.

DR GO; GO:0030374; F:ligand-dependent nuclear receptor transcrip. . . ; IDA.

DR GO; GO:0005515; F:protein binding; IPI.

SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 78.8%; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 2.7;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11

DB 168 EEIPLGKHYS 178

RESULT 2

Q46486 PRELIMINARY; PRT; 208 AA.

ID Q46486

AC Q46486;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein (Gcra).
 GN Gcra.
 OS Corynebacterium xerosis, and
 OS Corynebacterium striatum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1725, 43770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=C.xerosis; STRAIN=M82B;
 RX MEDLINE=96117603; PubMed=8559800;
 RA Tauch A., Kessing F., Kalinowski J., Puhler A.;
 RT "The 51,409-bp R-plasmid pTP10 from the multiresistant clinical
 RT isolate Corynebacterium striatum M82B is composed of DNA segments
 RT initially identified in soil bacteria and in plant, animal, and human
 RT pathogens";
 RL Mol. Gen. Genet. 263:1-11(2000).
 DR EMBL; U21300; AAC95478.1; -.
 DR EMBL; AF024666; AAC03390.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 208 AA; 23012 MW; F1504BE1CEDE85A6 CRC64;
 Query Match 69.2%; Score 36; DB 2; Length 208;
 Best Local Similarity 50.0%; Pred. No. 15;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPXGXHY 11
 Db 130 DVIPEGRHYA 139
 RESULT 3
 Q8CPJ4
 ID Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
 AC Q8CPJ4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain.
 GN SB0879.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE016746; AAC04476.1; -.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005524; F:AP binding; IEA.
 DR GO; GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0016874; F:ligase activity; IEA.
 DR GO; GO:0006526; P:arginine biosynthesis; IEA.
 DR GO; GO:0006807; P:nitrogen metabolism; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
 DR InterPro; IPR006275; CARA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.

DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS like.
 DR InterPro; IPR00169; SHProt acsite.
 DR Pfam; PF02389; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.
 DR Pfam; PF02787; CPase_L_D3; 1.
 DR Pfam; PF02142; MGS_L; 1.
 DR PRINTS; PRO0098; CPASE.
 DR TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
 DR PROSITE; PS00866; CPASE_1; 2.
 DR PROSITE; PS00867; CPASE_2; 2.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 KW Complete Proteome.
 SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;
 Query Match 69.2%; Score 35; DB 16; Length 1057;
 Best Local Similarity 63.6%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHY 11
 Db 189 KEVVSNGLHY 199
 RESULT 4
 Q40479
 ID Q40479 PRELIMINARY; PRT; 233 AA.
 AC Q40479;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE REBP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BY4; TISSUE=Leaf;
 RX MEDLINE=95276459; PubMed=7756828;
 RA Ohme-Takagi M., Shinshi H.;
 RT "ethylene-inducible DNA binding proteins that interact with an
 RT ethylene responsive element";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38126; BAA07324.1; -.
 DR PIR; T02590; T02590.
 DR HSP; O8037; 2GCC.
 DR TRANSFAC; T02654; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETRSPLEMMT.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2; 1.
 SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;
 Query Match 67.3%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 29;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHY 10
 Db 90 QAVVPKGRHY 99
 RESULT 5
 Q9LW50
 ID Q9LW50 PRELIMINARY; PRT; 237 AA.
 AC Q9LW50;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Ethylene-responsive element binding factor.
 GN NSERP2.
 OS Nicotiana sylvestris (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; asterids;
 OC lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20399450; PubMed=10945353;
 RA Kitajima S., Koyama T., Ohme-Takagi M., Shinshi H., Sato F.;
 RT "Characterization of gene expression of Nserfs, transcription factors
 of basic PR genes from Nicotiana sylvestris.";
 RL Plant Cell Physiol. 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1; -.
 DR HSSP; O80337; 2GCC.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR PRINTS; PR00367; ETRSPLEMT.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2; 1.
 DR SEQUENCE 237 AA; 26243 MW; 01BC3EB5E1E46298 CRC64;
 SQ
 Query Match 67.3%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 29;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEWVPGXHY 10
 DB 94 QAVVPGRHY 103
 RESULT 6
 ID Q98HU6 PRELIMINARY; PRT; 285 AA.
 AC Q98HU6;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Cytochrome c1.
 GN M12705.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=NAFF303099;
 RC MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 RA Kitanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.".
 RL Lact. Res. 7:331-338(2000).
 DR EMBL; AF003000; BAB4970.1; -.
 DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR InterPro; IPR002326; Cyt_C1.
 DR Pfam; PF02167; Cytochrome C1; 1.
 DR PRINTS; PR00603; CYTOCHROME_C1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 285 AA; 30961 MW; 31D9CDE2711747EE CRC64;

Query Match 67.3%; Score 35; DB 16; Length 285;
 Best Local Similarity 55.6%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXHY 11
 DB 194 VIPEGTHYN 202
 RESULT 7
 ID Q9X2E2 PRELIMINARY; PRT; 308 AA.
 AC Q9X2E2;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE PTH protease activity modulator HPLK.
 GN TM1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MS38 / DSM 3109;
 RC MEDLINE=92287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 DR EMBL; AE001819; AAD36885.1; -.
 DR PIR; A72207; A72207.
 DR TIGR; TM1822; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR InterPro; IPR001107; Band_7.
 DR InterPro; IPR001972; Stomatin.
 DR Pfam; PF01145; Band_7; 1.
 DR PRINTS; PR00721; STOMATIN.
 DR SMART; SM00244; PHB; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 308 AA; 34778 MW; ADE03603E5101A9D CRC64;
 Query Match 67.3%; Score 35; DB 16; Length 308;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXHY 10
 DB 41 VVPSGIHY 48
 RESULT 8
 ID Q38317 PRELIMINARY; PRT; 317 AA.
 AC Q38317;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Lysin.
 GN LYS.
 OS Lactobacillus bacteriophage phi adh.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=12417;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Altermann E.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=99384014; PubMed=10452953;
RA "Altermann E., Klein J., Heinrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RL Gasserii temperate bacteriophage phi-adh.";
RL Gene 236:333-346(1999).
[3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=95138034; PubMed=7836307;
RA Heinrich B., Binschofer B., Blaesi U.;
RT "Primary structure and functional analysis of the lysis genes of
RL Lactobacillus Gasserii bacteriophage phi-adh.";
RL J. Bacteriol. 177:723-732(1995).
[4]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=93231538; PubMed=8472961;
RA Pernaux C., De Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RL functions from Lactobacillus Gasserii temperate bacteriophage phi-
adh.";
RL Gene 126:61-66(1993).
[5]
RN RP SEQUENCE FROM N.A.
RX Engel G., Altermann E., Klein J., Heinrich B.;
RT "Structure of a genome region of the Lactobacillus Gasserii temperate
RL phase phi adh covering a repressor gene and cognate promoters.";
RL Gene 210:67-70(1998).
DR EMBL; AJ131519; CAB52540.1; -.
DR GO; GO:0003796; F:lysozyme activity; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002053; Glyco_hydro_25.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF01183; Glyco_hydro_25; 1.
DR ProDom; PD004620; Glyco_hydro_25; 1.
DR SMART; SM00641; Glyco_25; 1.
DR SMART; SM00287; SH3b; 1.
SQ SEQUENCE 317 AA; 34703 MW; 9FF2715EE43561C7 CRC64;

Query Match 67.3%; Score 35; DB 9; Length 317;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHYA 11
DB 60 VVPXGYHYA 68

RESULT 9
O27679 ID O27679 PRELIMINARY; PRT; 360 AA.
AC O27679;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cell division protein.
GN MTH1642.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Jamm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Fatwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

```

```

RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000923; AAB86115.1; -.
DR FNR; E69086; E69086.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0000910; P:cytokinesis; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR005140; eRF1_1.
DR InterPro; IPR005141; eRF1_2.
DR InterPro; IPR005142; eRF1_3.
DR InterPro; IPR004405; PclA.
DR Pfam; PF03463; eRF1_1; 1.
DR Pfam; PF03464; eRF1_2; 1.
DR Pfam; PF03465; eRF1_3; 1.
DR TIGRFAMs; TIGR00111; pelota; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 360 AA; 40814 MW; 2A000CB4B3CE469 CRC64;

Query Match 67.3%; Score 35; DB 17; Length 360;
Best Local Similarity 45.5%; Pred. No. 46;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYA 11
DB 98 EDLVPMGSHHT 108

RESULT 10
Q7XTG3 ID Q7XTG3 PRELIMINARY; PRT; 545 AA.
AC Q7XTG3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE QJ991214.12.4 protein.
GN QJ991214.12.4.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Wang Q.J., Zhang L., Lu Y.Q., Xu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL606453; CAB01515.1; -.
SQ SEQUENCE 545 AA; 59078 MW; 4629ASD1DA538692 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 545;
Best Local Similarity 55.6%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHYA 11
DB 433 IVPSGDHYA 441

RESULT 11
Q9ELX6 ID Q9ELX6 PRELIMINARY; PRT; 678 AA.
AC Q9ELX6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE Hypothetical protein.
 OS Cercopithecine herpesvirus 7.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=35245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
 RT "Complete sequence of the Simian Varicella Virus Genome."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275348; AAG27217.1; -
 DR GO; GO:0019012; C:vision; IEA.
 DR GO; GO:0006323; P:DNA packaging; IEA.
 DR InterPro; IPR007640; Herpes_UL17.
 DR Pfam; PF04559; Herpes_UL17; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 678 AA; 75850 MW; A17E09E30512FE3C CRC64;
 Query Match 67.3%; Score 35; DB 12; Length 678;
 Best Local Similarity 50.0%; Pred.No. 90;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGXHY 10
 Db 147 BEIIPKGRY 156
 ||:|:|:|
 RESULT 12
 Q8IME6 PRELIMINARY; PRT; 855 AA.
 ID Q8IME6;
 AC Q8IME6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG30437-PC.
 GN CG30437 OR CG10398 OR CG10408.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anatatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S. Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Brokstein P., Brokstein P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Fabrics B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoscin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Mazny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacieb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svingskæ R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weissenbach G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celnikner S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Anatatides P.G., Brandon R.C., Rogers Y.,
 RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacieb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikner S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.W.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celnikner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003786; AAM16124.1; -
 DR FlyBase; FBGN0050437; CG30437.
 DR GO; GO:0005507; F:copper ion binding; IEA.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR008972; Cupredoxin.
 DR InterPro; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
 SQ SEQUENCE 855 AA; 94532 MW; 39BD5A516D6312DB CRC64;
 Query Match 67.3%; Score 35; DB 5; Length 855;
 Best Local Similarity 66.7%; Pred.No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 BEVVPXGXH 9
 Db 657 DEVVPSGDH 665
 :|:|:|:|
 RESULT 13
 Q8YJ11 PRELIMINARY; PRT; 1028 AA.
 ID Q8YJ11
 AC Q8YJ11; 2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 25, Last annotation update)
 DE ATP-dependent DNA helicase.

```

GN BMEI0275.
OC Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
EX MEDLINE=20020109; PubMed=11756588;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonowski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyriades N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AE009470; AA51457.1; -.
DR FJL; AF3286; AF3286.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1028;
Best Local Similarity 54.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPXGXHS 11
Db 76 EKIVPPGARYS 86

RESULT 14
QSDIHO PRELIMINARY; PRT; 1044 AA.
AC QSDIHO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLI1618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
EX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shampo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130 (2002).
DR EMBL; AP005374; BAC09170.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR001036; Acrflavin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.

```

```

DR TIGRfams; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPXGXHS 11
Db 843 EEVLPMNGIGYS 853

RESULT 15
Q7WNB7 PRELIMINARY; PRT; 262 AA.
AC Q7WNB7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN B81123.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-598;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bason N., Cherevach I.,
RA Corden-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Doggett J.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moulé S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
RA Dawin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40 (2003).
DR EMBL; BX640440; CAE31621.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28907 MW; B3CA29331CB776B2 CRC64;

Query Match 65.4%; Score 34; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGXH 9
Db 182 QEVVPYQGH 190

Search completed: June 3, 2004, 11:57:33
Job time : 29.8667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 BEVVPXGXSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	93.8	11	5	ABB80549 Hepatitis
2	45	93.8	11	5	ABB80544 Hepatitis
3	45	93.8	11	5	ABB80553 Hepatitis
4	45	93.8	11	5	ABB80552 Hepatitis
5	45	93.8	11	5	ABB80545 Hepatitis
6	44	91.7	11	5	ABB80525 Hepatitis
7	44	91.7	11	5	ABB80521 Hepatitis
8	44	91.7	11	5	ABB80522 Hepatitis
9	44	91.7	11	5	ABB80536 Hepatitis
10	44	91.7	11	5	ABB80566 Hepatitis
11	44	91.7	11	5	ABB80563 Hepatitis
12	44	91.7	11	5	ABB80565 Hepatitis
13	44	91.7	11	5	ABB80535 Hepatitis
14	44	91.7	11	5	ABB80533 Hepatitis
15	44	91.7	11	5	ABB80567 Hepatitis
16	44	91.7	11	5	ABB80540 Hepatitis
17	44	91.7	11	5	ABB80559 Hepatitis
18	44	91.7	11	5	ABB80526 Hepatitis
19	44	91.7	11	5	ABB80530 Hepatitis
20	44	91.7	11	5	ABB80539 Hepatitis
21	44	91.7	11	5	ABB80564 Hepatitis
22	41	85.4	11	5	ABB80568 Hepatitis
23	41	85.4	11	5	ABB80548 Hepatitis
24	41	85.4	11	5	ABB80547 Hepatitis
25	41	85.4	11	5	ABB80556 Hepatitis
					ABB80557 Hepatitis

26 41 85.4 11 5 ABB80551 Hepatitis
27 40 83.3 11 5 ABB80534 Hepatitis
28 40 83.3 11 5 ABB80561 Hepatitis
29 40 83.3 11 5 ABB80542 Hepatitis
30 40 83.3 11 5 ABB80543 Hepatitis
31 40 83.3 11 5 ABB80546 Hepatitis
32 40 83.3 11 5 ABB80524 Hepatitis
33 40 83.3 11 5 ABB80533 Hepatitis
34 40 83.3 11 5 ABB80529 Hepatitis
35 40 83.3 11 5 ABB80528 Hepatitis
36 40 83.3 11 5 ABB80538 Hepatitis
37 40 83.3 11 5 ABB80562 Hepatitis
38 40 83.3 11 5 ABB80554 Hepatitis
39 40 83.3 11 5 ABB80550 Hepatitis
40 40 83.3 11 5 ABB80555 Hepatitis
41 39 81.2 11 5 ABB80523 Hepatitis
42 39 81.2 11 5 ABB80558 Hepatitis
43 39 81.2 11 5 ABB80537 Hepatitis
44 39 81.2 11 5 ABB80560 Hepatitis
45 39 81.2 11 5 ABB80527 Hepatitis

ALIGNMENTS

RESULT 1

ABB80549

ID ABB80549 standard; peptide; 11 AA.

XX ABB80549;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #29.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 5 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WO-US023169.

21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.

Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus

PT protease.

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

SO Sequence 11 AA;

Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EVVVPXGXSYS 11
 |||||
 Db 1 EVVVPXGSSYS 11

RESULT 2

ABB80544
 ID ABB80544 standard; peptide; 11 AA.

AC ABB80544;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #24.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

FN 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US023169.

PR 21-JUL-2000; 2000US-0220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-Wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

SO Sequence 11 AA;

Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 EVVVPXGXSYS 11
 |||||
 Db 1 EVVVPXGTSYS 11

RESULT 3

ABB80553
 ID ABB80553 standard; peptide; 11 AA.

AC ABB80553;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #33.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

FN 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US023169.

PR 21-JUL-2000; 2000US-0220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-Wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

SO Sequence 11 AA;

Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EVVVPXGXSYS 11
 |||||
 Db 1 EVVVPXGSSYS 11

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.
 DE
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 XX Synthetic.
 OS
 XX
 XX Location/Qualifiers
 FH Key
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PP
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 |||||
 DB 1 EEVVPXGTSYS 11
 |||||
 RESULT 4
 ABB80552
 ID ABB80552 standard; peptide; 11 AA.
 XX
 AC ABB80552;
 XX
 XX 08-OCT-2002 (first entry)
 DT
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #32.
 DE
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 XX Synthetic.
 OS
 XX
 XX Location/Qualifiers
 FH Key
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PP
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 |||||
 DB 1 EEVVPXGSSYS 11
 |||||
 RESULT 5
 ABB80545
 ID ABB80545 standard; peptide; 11 AA.
 XX
 AC ABB80545;
 XX
 XX 08-OCT-2002 (first entry)
 DT

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.
 DE
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 XX Synthetic.
 OS
 XX
 XX Location/Qualifiers
 FH Key
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PP
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 93.8%; Score 45; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.014;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 |||||
 DB 1 EEVVPXGTSYS 11
 |||||
 RESULT 6
 ABB80525
 ID ABB80525 standard; peptide; 11 AA.
 XX
 AC ABB80525;
 XX
 XX 08-OCT-2002 (first entry)
 DT
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
 DE
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 XX Synthetic.
 OS
 XX

PH	Key	Location/Qualifiers	
FT	Modified-site	1	
FT	FT	/note= "N-terminal acetyl"	
FT	Modified-site	6	
FT	FT	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"	
FT	Misc-difference	8	
FT	FT	/note= "D-form residue"	
FT	Modified-site	11	
FT	FT	/note= "C-terminal amide"	
XX	WO200208251-A2.		
XX	31-JAN-2002.		
XX	19-JUL-2001; 2001WO-US023169.		
XX	21-JUL-2000; 2000US-0220101P.		
XX	(CORV-) CORVAS INT INC.		
XX	Lim-Wilby M, Levy OE, Brunk TK;		
XX	WPI; 2002-361643/39.		
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.		
XX	Claim 17; Page 64; 69pp; English.		
XX	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus		
XX	Sequence 11 AA;		
XX	Query Match	91.7%; Score 44; DB 5; Length 11;	
XX	Best Local Similarity	90.9%; Pred. No. 0.023;	
XX	Matches	10; Conservative 0; Mismatches 1; Indels 0; Gaps 0	
Qy	1 EEVFPXGXSYS 11		
Db	1 EEVFPXGXSYS 11		
RESULT 7			
ABB80521			
ID	ABB80521 standard; peptide; 11 AA.		
XX	AC		
XX	ABB80521;		
XX	08-OCT-2002 (first entry)		
XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.		
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.		
XX	Synthetic.		
XX	Key	Location/Qualifiers	
XX	Modified-site	1	
XX	FT	/note= "N-terminal acetyl"	
XX	Modified-site	6	
XX	FT	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"	
XX	Modified-site	11	
XX	FT	/note= "C-terminal amide"	

XX (CORV-) CORVAS INT INC.
 XX Lim-Wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 XX Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 XX Sequence 11 AA;
 SQ
 Query Match 91.7%; Score 44; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 DB 1 EEVVPXGXSYS 11
 RESULT 9
 ABB80536
 ID ABB80536 standard; peptide; 11 AA.
 XX
 AC ABB80536;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #16.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 XX Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 XX Sequence 11 AA;
 SQ
 Query Match 91.7%; Score 44; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 DB 1 EEVVPXGXSYS 11
 RESULT 10
 ABB80566
 ID ABB80566 standard; peptide; 11 AA.
 XX
 AC ABB80566;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "2-aminoisobutyryl carbonyl residue forming a keto
 FT -amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 XX
 PN WO200208251-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US023169.
 XX
 PR 21-JUL-2000; 2000US-0220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-Wilby M, Levy OE, Brunck TK;
 XX
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX
 XX Claim 17; Page 65; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 91.7%; Score 44; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.023;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11

RESULT 11

ABB80563
 ID ABB80563 standard; peptide; 11 AA.

XX AC ABB80563;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

XX OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Valyl carbonyl forming keto-amide linkage with
 residue 7"

FT Modified-site 11

FT /note= "C-terminal amide"

XX WO200208251-A2.

PN 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX Sequence 11 AA;

Query Match 91.7%; Score 44; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.023;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11

Db 1 EEVVPXGMSYS 11
 |||||

RESULT 12

ABB80565
 ID ABB80565 standard; peptide; 11 AA.

XX AC ABB80565;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

XX OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norleucyl carbonyl forming keto-amide linkage
 with residue 7"

FT Modified-site 11

FT /note= "C-terminal amide"

XX WO200208251-A2.

PN 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

XX Sequence 11 AA;

Query Match 91.7%; Score 44; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.023; 1; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11

Db 1 EEVVPXGMSYS 11

RESULT 13

ABB80535
 ID ABB80535 standard; peptide; 11 AA.

XX AC ABB80535;

XX


```

FT      /note= "C-terminal amide"
XX
FN      WO200208251-A2.
XX
PD      31-JAN-2002.
XX
PF      19-JUL-2001; 2001WO-US023169.
XX
PR      21-JUL-2000; 2000US-0220101P.
XX
PA      (CORV-) CORVAS INT INC.
XX
PI      Lim-Wilby M, Levy OE, Brunck TK;
XX
DR      WPI; 2002-361643/39.
XX
PT      Novel peptide compound having hepatitis C virus protease inhibitory
PT      activity useful for treating disorders associated with hepatitis C virus
PT      protease.
XX
PS      Claim 17; Page 65; 69pp; English.
XX
CC      The sequence represents a peptide compound of the invention having
CC      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC      invention are alpha-ketoamide peptide analogues. The peptides have
CC      virucide activity, and are useful for treating and in the manufacture of
CC      a medicament to treat disorders associated with HCV protease. A
CC      pharmaceutical composition comprising the peptide as an active ingredient
CC      is useful for treating disorders associated with hepatitis C virus
XX
SQ      Sequence 11 AA;

Query Match          91.7%; Score 44; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.023;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGXSYS 11
        |||||
DB      1 EEVVPXGQSYS 11

Search completed: June 3, 2004, 11:48:24
Job time : 45.9333 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-43
Perfect score: 48
Sequence: 1 BEVVPXGXSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	36	75.0	4 US-09-408-020-4	Sequence 4, Appli
2	32	66.7	45 US-08-637-759B-236	Sequence 236, App
3	32	66.7	45 US-08-871-355A-236	Sequence 236, App
4	32	66.7	45 US-09-201-945-236	Sequence 236, App
5	32	66.7	947 US-09-228-986-73	Sequence 73, Appl
6	31	64.6	159 US-08-844-086-4	Sequence 4, Appli
7	31	64.6	159 US-09-018-211-4	Sequence 4, Appli
8	31	64.6	181 US-09-134-000C-4848	Sequence 4848, Ap
9	31	64.6	507 US-09-424-978B-34	Sequence 34, Appl
10	31	64.6	513 PCT-US91-02714-26	Sequence 26, Appl
11	31	64.6	622 US-08-459-146-2	Sequence 2, Appli
12	31	64.6	622 US-08-459-065-2	Sequence 2, Appli
13	31	64.6	688 1 US-07-688-352C-28	Sequence 28, Appl
14	31	64.6	688 2 US-08-474-379C-28	Sequence 28, Appl
15	31	64.6	688 3 US-09-146-249A-28	Sequence 28, Appl
16	31	64.6	688 3 US-08-206-188B-28	Sequence 28, Appl
17	31	64.6	833 3 US-08-844-086-2	Sequence 2, Appli
18	31	64.6	833 3 US-09-018-211-2	Sequence 2, Appli
19	31	64.6	1407 4 US-09-328-352-7885	Sequence 7885, Ap
20	30	62.5	121 4 US-09-153-060-68	Sequence 68, Appl
21	30	62.5	121 4 US-09-153-060-85	Sequence 85, Appl
22	30	62.5	122 2 US-08-879-985A-1	Sequence 1, Appli
23	30	62.5	122 3 US-09-213-096-1	Sequence 1, Appli
24	30	62.5	241 3 US-08-834-776A-2	Sequence 2, Appli
25	30	62.5	382 4 US-09-134-000C-3738	Sequence 3738, Ap
26	30	62.5	404 4 US-09-498-520A-42	Sequence 42, Appl
27	30	62.5	480 1 US-07-803-636A-2	Sequence 2, Appli

Sequence 220, App
Sequence 220, App
Sequence 6588, Ap
Sequence 93, Appl
Sequence 80, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 53, Appl
Sequence 8, Appli
Sequence 8, Appli
Sequence 27, Appl
Sequence 27, Appl
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli

US-08-961-083-220
US-09-536-784-220
US-09-134-000C-6588
US-09-413-814-93
US-09-413-814-80
US-08-361-517-19
PCT-US93-07964-19
US-08-440-861-53
US-07-971-096-8
US-08-175-096-8
US-09-047-125-27
US-07-736-335B-27
US-08-413-974-4
US-08-434-418-4
US-08-433-288-4
US-08-174-739A-4
US-08-434-256-4
US-07-971-096-4

ALIGNMENTS

RESULT 1

US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632337
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCOIP 002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 75.0%; Score 36; DB 4; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.8e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11

DB 2294 EDVTPRGISFS 2304

RESULT 2

US-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 66.7%; Score 32; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 9.9;
Matches 6; Conservative 1; Mismatches 3; Indels 0;
Gaps 0;

QY 1 EEVVPXGXSY 10
||: ||: ||
Db 1 EEISPLGWSY 10

RESULT 3
US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:

SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 66.7%; Score 32; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 9.9;
Matches 5; Conservative 1; Mismatches 3; Indels 0;
Gaps 0;

QY 1 EEVVPXGXSY 10
||: ||: ||
Db 1 EEISPLGWSY 10

RESULT 4
US-09-201-945-236
Sequence 236, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-236

Query Match 66.7%; Score 32; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 9.9;
Matches 6; Conservative 1; Mismatches 3; Indels 0;
Gaps 0;

QY 1 EEVVPXGXSY 10
||: ||: ||
Db 1 EEISPLGWSY 10

RESULT 5
US-09-228-986-73
Sequence 73, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11006/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-09-228-986-73

Query Match 66.7%; Score 32; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

3 VVPXGXSYS 11
686 VMPXGXSYS 694

RESULT 6
US-08-844-086-4
Sequence 4, Application US/08844086
Patent No. 5866390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390e1 Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: F31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-086-4

Query Match 64.6%; Score 31; DB 2; Length 159;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 EVVPXGXS 9
123 EVVLPDGT 131

RESULT 7
US-09-018-211-4
Sequence 4, Application US/09018211
Patent No. 6048716
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6048716e1 Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,211
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,086
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-018-211-4

Query Match 64.6%; Score 31; DB 3; Length 159;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 EVVPXGXS 9
123 EVVLPDGT 131

RESULT 8
US-09-134-000C-4848
Sequence 4848, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032

;; CURRENT APPLICATION NUMBER: US/09/134,000C
;; CURRENT FILING DATE: 1998-08-13
;; PRIOR APPLICATION NUMBER: US 60/055,778
;; PRIOR FILING DATE: 1997-08-15
;; NUMBER OF SEQ ID NOS: 6812
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 4848
;; LENGTH: 181
;; TYPE: PRT
;; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 64.6%; Score 31; DB 4; Length 181;
Best Local Similarity 60.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 EVVVPXGXS Y 10
|||
Db 145 EVVPTSEDY 154

RESULT 9
US-09-424-978B-34

;; Sequence 34, Application US/09424978B
;; Patent No. 6664445

;; GENERAL INFORMATION:

;; APPLICANT: Falco, Saverio Carl

;; APPLICANT: Allen, Stephen M.

;; APPLICANT: Rafalski J. Antoni

;; APPLICANT: Hitz, William D.

;; APPLICANT: Kinney, Anthony J.

;; APPLICANT: Abell, Lynne N.

;; APPLICANT: Thorpe, Catherine J.

;; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

;; FILE REFERENCE: BB-1087

;; CURRENT APPLICATION NUMBER: US/09/424,978B

;; CURRENT FILING DATE: 1999-12-02

;; PRIOR APPLICATION NUMBER: US 60/048,771

;; PRIOR FILING DATE: 1997-06-06

;; NUMBER OF SEQ ID NOS: 43

;; SOFTWARE: Patent in version 3.1

;; SEQ ID NO 34

;; LENGTH: 507

;; TYPE: PRT

;; ORGANISM: Burkholderia capacia

US-09-424-978B-34

Query Match

Best Local Similarity 64.6%; Score 31; DB 4; Length 507;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVVPXGXS Y 11

|||

Db 120 EVIQAGESYS 129

RESULT 10

PCT-US91-02714-26

;; Sequence 26, Application PC/TUS9102714

;; GENERAL INFORMATION:

;; APPLICANT: Wigler, Michael H.

;; APPLICANT: Colicelli, John J.

;; TITLE OF INVENTION: Cloning by Complementation and Related

;; TITLE OF INVENTION: Processes

;; NUMBER OF SEQUENCES: 55

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

;; ADDRESSEE: Bicknell

;; STREET: Two First National Plaza, 20 South Clark

;; STREET: Street

;; CITY: Chicago

;; STATE: Illinois

;; COUNTRY: USA

;; ZIP: 60603
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US91/02714
;; FILING DATE: 19910419
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/511,715
;; FILING DATE: 20-APR-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Borun, Michael F.
;; REGISTRATION NUMBER: 25447
;; REFERENCE/DOCKET NUMBER: 27805/30197
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 346-5750
;; TELEFAX: (312) 984-9740
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 26:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 513 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US91-02714-26

Query Match 64.6%; Score 31; DB 5; Length 513;

Best Local Similarity 75.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVVPXGXS Y 10

|||

Db 201 VVPAGSY 208

RESULT 11

US-08-459-146-2

;; Sequence 2, Application US/08459146

;; Patent No. 5866405

;; GENERAL INFORMATION:

;; APPLICANT: Choi, Gil Ho

;; APPLICANT: Nuss, Donald Lee

;; TITLE OF INVENTION: Genetically Engineered Transmissible

;; NUMBER OF SEQUENCES: 3

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.

;; STREET: 340 Kingsland Street

;; CITY: Nutley

;; STATE: New Jersey

;; COUNTRY: U.S.A.

;; ZIP: 07110

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent in Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/459,146

;; FILING DATE: 02-JUN-1995

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/832,117

;; FILING DATE: 06-FEB-1992

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Roseman, Catherine R

;; REGISTRATION NUMBER: 34,240

;; REFERENCE/DOCKET NUMBER: 8589

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (201) 235-6208

TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryptonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-146-2

Query Match 64.6%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXG 7
DB 31 EVVVPAG 37

RESULT 12
US-08-459-065-2
Sequence 2, Application US/08459065
Patent No. 5882642
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryptonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-065-2

Query Match 64.6%; Score 31; DB 2; Length 622;

Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EVVVPXG 7
DB 31 EVVVPAG 37

RESULT 13
US-07-688-352C-28
Sequence 28, Application US/07688352C
Patent No. 5527896
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-688-352C-28

Query Match 64.6%; Score 31; DB 1; Length 688;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSX 10
DB 201 VVPAGGSY 208

RESULT 14
US-08-474-379C-28
Sequence 28, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
TITLE OF INVENTION: PROCESSES

NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/474,379C
APPLICATION NUMBER: US/08/474,379C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-474-379C-28

Query Match 64.6%; Score 31; DB 2; Length 688;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXSY 10
||| |||
Db 201 VVPAGGSY 208

RESULT 15
US-09-146-249A-28
Sequence 28, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-146-249A-28

Query Match 64.6%; Score 31; DB 3; Length 688;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXSY 10
||| |||
Db 201 VVPAGGSY 208

Search completed: June 3, 2004, 12:03:08
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)

91.741 Million cell updates/sec

Title: US-09-909-164-43
Perfect score: 48
Sequence: 1 EEVVPXGXSYS 11

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	93.8	11	12	US-09-909-164-28
2	45	93.8	11	12	US-09-909-164-29
3	45	93.8	11	12	US-09-909-164-33
4	45	93.8	11	12	US-09-909-164-36
5	45	93.8	11	12	US-09-909-164-37
6	44	91.7	11	12	US-09-909-164-5
7	44	91.7	11	12	US-09-909-164-6
8	44	91.7	11	12	US-09-909-164-9
9	44	91.7	11	12	US-09-909-164-10
10	44	91.7	11	12	US-09-909-164-14
11	44	91.7	11	12	US-09-909-164-19
12	44	91.7	11	12	US-09-909-164-20
13	44	91.7	11	12	US-09-909-164-23
14	44	91.7	11	12	US-09-909-164-24
15	44	91.7	11	12	US-09-909-164-43

16	44	91.7	11	12	US-09-909-164-47
17	44	91.7	11	12	US-09-909-164-48
18	44	91.7	11	12	US-09-909-164-49
19	44	91.7	11	12	US-09-909-164-50
20	44	91.7	11	12	US-09-909-164-51
21	44	91.7	11	12	US-09-909-164-52
22	41	85.4	11	12	US-09-909-164-31
23	41	85.4	11	12	US-09-909-164-32
24	41	85.4	11	12	US-09-909-164-35
25	41	85.4	11	12	US-09-909-164-40
26	41	85.4	11	12	US-09-909-164-41
27	40	83.3	11	12	US-09-909-164-8
28	40	83.3	11	12	US-09-909-164-12
29	40	83.3	11	12	US-09-909-164-13
30	40	83.3	11	12	US-09-909-164-17
31	40	83.3	11	12	US-09-909-164-18
32	40	83.3	11	12	US-09-909-164-22
33	40	83.3	11	12	US-09-909-164-26
34	40	83.3	11	12	US-09-909-164-27
35	40	83.3	11	12	US-09-909-164-30
36	40	83.3	11	12	US-09-909-164-34
37	40	83.3	11	12	US-09-909-164-38
38	40	83.3	11	12	US-09-909-164-39
39	40	83.3	11	12	US-09-909-164-45
40	40	83.3	11	12	US-09-909-164-46
41	39	81.2	11	12	US-09-909-164-7
42	39	81.2	11	12	US-09-909-164-11
43	39	81.2	11	12	US-09-909-164-15
44	39	81.2	11	12	US-09-909-164-16
45	39	81.2	11	12	US-09-909-164-21

ALIGNMENTS

RESULT 1

US-09-909-164-28
; Sequence 28, Application US/0909164
; Publication No: US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunch, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)-(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)-(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-28

Query Match 93.8% ; Score 45; DB 12; Length 11;

OTHER INFORMATION: D-amino acid
US-09-909-164-36

Query Match 93.8%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0077; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXSYS 11
| | | | | | | | | | |
DB 1 EVVVPXGSSYS 11

RESULT 5

US-09-909-164-37
Sequence 37, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patentin version 3.1
SEQ ID NO 37
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLTATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-37

Query Match 93.8%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0077;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXSYS 11
| | | | | | | | | | |
DB 1 EVVVPXGSSYS 11

RESULT 6

US-09-909-164-5
Sequence 5, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patentin version 3.1
SEQ ID NO 5
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLTATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-5

Query Match 91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXSYS 11
| | | | | | | | | | |
DB 1 EVVVPXGMSYS 11

RESULT 7

US-09-909-164-6
Sequence 6, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Brunnck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patentin version 3.1
SEQ ID NO 6
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLTATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)

OTHER INFORMATION: D-amino acid
NAME/KEY: MOD_RES
LOCATION: (11)..(11)

```
; OTHER INFORMATION: AMIDATION
US-09-909-164-6
Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy 1 EEVVPXGXSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 8
US-09-909-164-9
; Sequence 9, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; OTHER INFORMATION: 164-10
US-09-909-164-10
Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy 1 EEVVPXGXSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 10
US-09-909-164-14
; Sequence 14, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 14
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: 164-9
US-09-909-164-9
Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy 1 EEVVPXGXSYS 11
Db 1 EEVVPXGMSYS 11

RESULT 9
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
```

```
; OTHER INFORMATION: norvaline-(CO)
JS-09-909-164-14

Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
Db 1 BEVVPXGGSYS 11

RESULT 11
US-09-909-164-19
; Sequence 19, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; OTHER INFORMATION: 11-mer synthesized according to example 1
US-09-909-164-20

Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
Db 1 BEVVPXGGSYS 11

RESULT 13
US-09-909-164-23
; Sequence 23, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
```



```
; OTHER INFORMATION: D-amino acid
US-09-909-164-23
Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSYS 11
   |||||
Db 1 EEVVPXGXSYS 11

RESULT 14
US-09-909-164-24
; Sequence 24, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-24

Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSYS 11
   |||||
Db 1 EEVVPXGXSYS 11

RESULT 15
US-09-909-164-43
; Sequence 43, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
```

```
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-43

Query Match          91.7%; Score 44; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSYS 11
   |||||
Db 1 EEVVPXGXSYS 11

Search completed: June 3, 2004, 12:57:16
Job time : 33.7333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

PM protein - protein search, using sw model

run on: June 3, 2004, 11:35:47 ; Search time 9 seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 EEVVPXGXSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : FIR 78.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	36	75.0	3472	T31308	hypothetical 367K
2	35	72.9	225	S57810	hypothetical prote
3	34	70.8	102	A42452	V1 protein - tobac
4	34	70.8	165	D69493	hypothetical prote
5	34	70.8	259	T34536	hypothetical prote
6	34	70.8	1028	AF3286	ATP-dependent DNA
7	33	68.8	124	VKLJ51	trans-regulatory s
8	33	68.8	427	F64064	tolB protein - Hae
9	32	66.7	227	E75619	hypothetical prote
10	32	66.7	425	T24111	hypothetical prote
11	32	66.7	426	D82163	3-phosphohikimate
12	32	66.7	670	S22293	zinc finger protei
13	32	66.7	890	A30481	bacteriocin BCN5
14	32	66.7	2717	A34203	DNA-binding protei
15	31	64.6	123	B69342	conserved hypothet
16	31	64.6	284	S75817	hypothetical prote
17	31	64.6	319	S03833	hypothetical prote
18	31	64.6	361	S15299	dTDP-glucose 4,6-de
19	31	64.6	361	AF0767	hypothetical prote
20	31	64.6	437	AG2945	periplasmic sorbit
21	31	64.6	450	C98337	ATP-dependent DNA
22	31	64.6	541	AH2679	probable ABC subst
23	31	64.6	544	C82900	type II secretion
24	31	64.6	561	C84239	DNA ligase (AB0425
25	31	64.6	573	P97461	hypothetical prote
26	31	64.6	612	T05331	hypothetical prote
27	31	64.6	622	S15009	probable beta-gala
28	31	64.6	646	C95978	iron(III) ABC tran
29	31	64.6	653	D82352	

30	31	64.6	701	2	S61239	hypothetical prote
31	31	64.6	829	2	H86726	leucine-trna ligas
32	31	64.6	833	2	H95029	leucyl-trna synthe
33	31	64.6	833	2	C97901	leucine-trna ligas
34	31	64.6	840	2	T39116	probable sulfate p
35	31	64.6	877	2	T40413	sulfate permease -
36	31	64.6	926	2	AG1860	hypothetical prote
37	31	64.6	1014	2	T17275	hypothetical prote
38	31	64.6	1081	1	A42399	isoleucine-trna li
39	31	64.6	1085	2	T18369	K-Cl cotransport p
40	31	64.6	1085	2	T13429	K-Cl cotransport p
41	31	64.6	1086	2	T14114	K-Cl cotransport p
42	31	64.6	1088	2	D82246	probable chitinase
43	31	64.6	1116	2	T31432	K-Cl cotransport p
44	31	64.6	1152	2	D87046	conserved hypothet
45	31	64.6	1548	2	T04456	hypothetical prote

ALIGNMENTS

RESULT 1

T31308

hypothetical 367K protein - Cenarchaeum symbiosum

C;Species: Cenarchaeum symbiosum

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000

C;Accession: T31308

R;Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.

J. Bacteriol. 180, 5003-5009, 1998

A;Title: Genomic analysis reveals chromosomal variation in natural populations of the ur

A;Reference number: Z20994; MUID:98422450; PMID:9749430

A;Accession: T31308

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3472 <SCH>

A;Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC62699.1

C;Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 75.0%; Score 36; DB 2; Length 3472;

Best Local Similarity 54.5%; Pred. No. 79;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSYS 11

Db 2294 EDVPRGISFS 2304

RESULT 2

S57810

hypothetical protein precursor (clone Tpp11) - tomato

C;Species: Lycopersicon esculentum (tomato)

C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000

C;Accession: S57810

R;Milligan, S.B.; Gasser, C.S.

Plant Mol. Biol. 28, 691-711, 1995

A;Title: Nature and regulation of plastid-expressed genes in tomato.

A;Reference number: S57808; MUID:95375233; PMID:7647301

A;Accession: S57810

A;Status: preliminary; nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-225 <MIL>

A;Cross-references: EMBL:U20592; NID:G924625; PIDN:AAA80497.1; PID:G924626

C;Superfamily: plant Kunitz-type proteinase inhibitor

Query Match

Best Local Similarity 72.9%; Score 35; DB 2; Length 225;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSYS 11

Db 32 DEVVPGKTYA 42

```

RESULT 3
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C;Species: tobacco yellow dwarf virus
C;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C;Accession: A42452
R;Morris, B.A.M.; Richardson, K.A.; Halsey, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A;Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A;Reference number: A42452; MUID:92188538; PMID:1546458
A;Accession: A42452
A;Molecule type: DNA
A;Residues: 1-102 <MOR>
A;Cross-references: GB:M81103; NID:9335283; PIDN:AAA47947.1; PID:g335284

Query Match 70.8%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 4.9;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGXSYS 11
Db 7 QVVPFSGINYS 16

RESULT 4
D69493
hypothetical protein AF1949 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 04-Mar-2000
C;Accession: D69493
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Globe, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Ueberback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: D69493
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-165 <KLE>
A;Cross-references: GB:AE000968; GB:AE000782; NID:92689291; PIDN:AA89307.1; PID:g264859
C;Superfamily: Archaeoglobus fulgidus hypothetical protein AF1949

Query Match 70.8%; Score 34; DB 2; Length 165;
Best Local Similarity 60.0%; Pred. No. 8.3;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGXSYS 10
Db 60 EESIPDGASY 69

RESULT 5
T34536
hypothetical protein DKFZp34C031.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34536
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21540
A;Accession: T34536
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-259 <POU>
A;Cross-references: EMBL:AL122063
A;Experimental source: adult testis; clone DKFZp34C031
C;Genetics:
A;Note: DKFZp34C031.1

Query Match 70.8%; Score 34; DB 2; Length 259;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Best Local Similarity 60.0%; Pred. NO. 13;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGXSYS 11
Db 22 EVAPAGASYN 31

RESULT 6
AF3286
ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AF3286
R;Pellicchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagijs, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3286
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1028 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51457.1; PID:g17982167; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0275
A;Map position: 1

Query Match 70.8%; Score 34; DB 2; Length 1028;
Best Local Similarity 54.5%; Pred. NO. 59;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
Db 76 EKIVPFGARIS 86

RESULT 7
VKLJ51
trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz
N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs prote
C;Species: simian immunodeficiency virus SIVcpz
A;Note: host Pan troglodytes (chimpanzee)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C;Accession: S09988
R;Ruet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
Nature 345, 356-359, 1990
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
A;Reference number: S09983; MUID:90259077; PMID:2188136
A;Accession: S09988
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-124 <HUS>
A;Cross-references: EMBL:X52154; NID:958866; PIDN:CAA36405.1; PID:g763085
C;Genetics:
A;Gene: rev; trs; art
A;Introns: 27/1
C;Superfamily: AIDS trans-regulatory splicing protein
C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 68.8%; Score 33; DB 1; Length 124;
Best Local Similarity 60.0%; Pred. NO. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGXSYS 11
Db 107 ETVPAGGINS 116

RESULT 8
F64064
colb protein - Haemophilus influenzae (strain Rd KW20)

```


A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-670 <NT>

A:Cross-references: EMBL:X54250; NID:G57519; PIDN:CAA38151.1; PID:G57520

A>Note: the authors did not translate the codon for residue 1

C:Superfamily: HIV-EP2 enhancer-binding protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 66.7%; Score 32; DB 2; Length 670;

Best Local Similarity 66.7%; Pred. No. 1e+02; Indels 2; Gaps 0;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSYS 11

DB 376 VVPAGLTYS 384

RESULT 13

A30481

bacteriocin BCN5 - Clostridium perfringens plasmid pIP404

C:Species: Clostridium perfringens

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Nov-1999

C:Accession: A30481; S03779

R:Garnier, T.; Cole, S.T.

J. Bacteriol. 168, 1189-1196, 1986

A:Title: Characterization of a bacteriocinogenic plasmid from Clostridium perfringens

A:Reference number: J0354; MUID:87057020; PMID:2877971

A:Accession: A30481

A:Molecule type: DNA

A:Residues: 1-890 <GAR>

A:Cross-references: GB:M32882; GB:J03309; NID:G150738; PIDN:AAA98249.1; PID:G150739

C:Genetics:

A:Gene: bcn

A:Genome: plasmid

C:Superfamily: Clostridium perfringens plasmid pIP404 bacteriocin BCN5

C:Keywords: bacteriocin

Query Match

Best Local Similarity 66.7%; Score 32; DB 2; Length 890;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSY 10

DB 170 EVVPGGFTY 178

RESULT 14

A34203

DNA-binding protein PRDII-BF1 - human

N:Alternate names: major histocompatibility complex enhancer-binding protein 1

C:Species: Homo sapiens (man)

C>Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999

C:Accession: A34203; A34779

R:Fan, C.M.; Maniatis, T.

Genes Dev. 4, 29-42, 1990

A:Title: A DNA-binding protein containing two widely separated zinc finger motifs that

A:Reference number: A34203; MUID:90169514; PMID:2106471

A:Accession: A34203

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2717 <FAN>

A:Cross-references: EMBL:X51435; NID:G38017; PIDN:CAA35798.1; PID:G38018

R:Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.

Mol. Cell. Biol. 10, 1406-1414, 1990

A:Title: A large protein containing zinc finger domains binds to related sequence element

A:Reference number: A34779; MUID:90205817; PMID:2108316

A:Accession: A34779

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-16

A:Cross-references: GB:M32019

C:Superfamily: HIV-EP2 enhancer-binding protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 66.7%; Score 32; DB 2; Length 2717;

Best Local Similarity 66.7%; Pred. No. 4.5e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSYS 11

DB 2405 VVPAGLTYS 2413

RESULT 15

E89342

conserved hypothetical protein AF0741 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Dec-2002

C:Accession: E89342

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F

Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Sprigge, T.; Artiach, P.; Kaine, B.P.; Sykes, S

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: E89342

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-123 <KLE>

A:Cross-references: GB:AF001054; GB:AE000782; NID:G2689377; PIDN:AAB90501.1; PID:G26498

C:Superfamily: uncharacterized conserved protein

Query Match

Best Local Similarity 50.0%; Score 31; DB 2; Length 123;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSY 10

DB 51 EKIAPYGDSY 60

Search completed: June 3, 2004, 12:00:02

Job time : 10 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments) updates/sec
117.693 Million cell

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 EEVVPXGXSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	75.0	1499	1 A10A_HUMAN	O60312 homo sapien
2	35	72.9	1058	1 CARB_FUSNN	Q8rg86 fuscobacteri
3	34	70.8	102	1 V11K_TYDVA	P31619 tobacco yel
4	34	70.8	165	1 X749_ARCFU	O28330 archaeoglob
5	34	70.8	478	1 GSR2_HUMAN	Q9nm55 homo sapien
6	33	68.8	124	1 REV_SIVCZ	P17280 chimpanzee
7	33	68.8	427	1 TOLB_HAEIN	P44677 haemophilus
8	32	66.7	426	1 AROA_VIBCH	Q9krb0 vibrio chol
9	32	66.7	890	1 BCN5_CLOPE	P08696 clostridium
10	32	66.7	2717	1 ZEPI_HUMAN	P15822 homo sapien
11	31	64.6	319	1 VYAN_CRYPA	P10941 cryphonectr
12	31	64.6	361	1 RFBP_SALTY	P26391 salmonella
13	31	64.6	507	1 THDI_BURCE	P53607 burkholderi
14	31	64.6	829	1 SYL_LACIA	Q9chb6 lactococcus
15	31	64.6	833	1 SYL_STR3	O8e2v2 streptococc
16	31	64.6	833	1 SYL_STRMU	O8d8s5 streptococc
17	31	64.6	833	1 SYL_STRP3	O8k8e1 streptococc
18	31	64.6	833	1 SYL_STRP8	O8p2t2 streptococc
19	31	64.6	833	1 SYL_STRN	Q978s0 streptococc
20	31	64.6	833	1 SYL_STRY	Q9alp0 streptococc
21	31	64.6	833	1 SYL_STR36	O8drb6 streptococc
22	31	64.6	877	1 SULH_SCHPO	O74377 schizosacch
23	31	64.6	1081	1 SYL_TETTH	P36422 tetrahymena
24	31	64.6	1083	1 S127_MOUSE	Q9y666 homo sapien
25	31	64.6	1083	1 S127_MOUSE	Q9w13 mus musculus
26	31	64.6	1085	1 S124_HUMAN	Q9up95 homo sapien
27	31	64.6	1085	1 S124_MOUSE	Q9j1s8 mus musculus
28	31	64.6	1085	1 S124_RABIT	Q28677 oryctolagus
29	31	64.6	1085	1 S124_RAT	O63632 rattus norv
30	31	64.6	1115	1 S125_MOUSE	Q91v14 mus musculus
31	31	64.6	1116	1 S125_HUMAN	Q9h2x9 homo sapien
32	31	64.6	1116	1 S125_RAT	O63633 rattus norv
33	31	64.6	1150	1 S126_HUMAN	Q9uhw9 homo sapien

34	31	64.6	1150	1 S126_MOUSE	Q924n4 mus musculus
35	30	62.5	121	1 TKNK_HUMAN	Q9uhf0 homo sapien
36	30	62.5	223	1 PURQ_CORGL	Q8nm14 corynebacte
37	30	62.5	223	1 PURQ_PYRFO	Q8u492 pyrococcus
38	30	62.5	223	1 PURQ_PYRHO	O59e19 pyrococcus
39	30	62.5	224	1 PURQ_HALNI	Q9hnu2 halobacteri
40	30	62.5	224	1 PURQ_MYCLE	O05756 mycobacteri
41	30	62.5	224	1 PURQ_MYCTU	P71841 mycobacteri
42	30	62.5	225	1 PURQ_CORAM	Q9rbx0 corynebacte
43	30	62.5	232	1 PURQ_METAC	Q8tpf0 methanosarc
44	30	62.5	232	1 PURQ_METMA	Q8ftb0 methanosarc
45	30	62.5	232	1 SCOA_HELPJU	Q92ie3 helicobacte

ALIGNMENTS

RESULT 1
A10A_HUMAN STANDARD; PRT; 1499 AA.
ID AC O60312; Q96914;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
DE (Aminophospholipid translocase VA).
GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2125279; PubMed=11326269;
RA Meguro M., Kashiwagi A., Mitsuya K., Nakao M., Kondo I., Saichou S.,
RA Oshimura M.;
RT "A novel maternally expressed gene, ATP10C, encodes a putative
RT aminophospholipid translocase associated with Angelman syndrome.";
RL Nat. Genet. 28:19-20(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21313119; PubMed=11353404;
RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
RT "The human aminophospholipid-transporting ATPase gene ATP10C maps
RT adjacent to UBR3A and exhibits similar imprinted expression.";
RL Am. J. Hum. Genet. 68:1501-1505(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Skin;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny P.J., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 337-1499 FROM N.A.
RN TISSUE=Brain;

RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- TISSUE SPECIFICITY: Widely expressed, with highest levels in
CC kidney, followed by lung, brain, prostate, testis, ovary and
CC small intestine.
CC -1- DISEASE: Defects in ATP10A are a cause of Angelman syndrome (AS)
CC [MIM:105830]; also known as 'happy puppet syndrome'. AS is
CC characterized by features of severe motor and intellectual
CC retardation, microcephaly, ataxia, frequent jerky limb movements
CC and flapping of the arms and hands, hypotonia, hyperactivity,
CC hypopigmentation, seizures, absence of speech, frequent smiling
CC and episodes of paroxysmal laughter, and an unusual facies
CC characterized by macrosomia, a large mandible and open-mouthed
CC expression, a great propensity for protruding the tongue ('tongue
CC thrusting'), and an occipital groove.
CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB051358; BAB47392.1; -
DR EMBL; AY029504; AAK33100.1; JOINED.
DR EMBL; AY029487; AAK33100.1; JOINED.
DR EMBL; AY029488; AAK33100.1; JOINED.
DR EMBL; AY029489; AAK33100.1; JOINED.
DR EMBL; AY029490; AAK33100.1; JOINED.
DR EMBL; AY029491; AAK33100.1; JOINED.
DR EMBL; AY029492; AAK33100.1; JOINED.
DR EMBL; AY029493; AAK33100.1; JOINED.
DR EMBL; AY029494; AAK33100.1; JOINED.
DR EMBL; AY029495; AAK33100.1; JOINED.
DR EMBL; AY029496; AAK33100.1; JOINED.
DR EMBL; AY029497; AAK33100.1; JOINED.
DR EMBL; AY029498; AAK33100.1; JOINED.
DR EMBL; AY029499; AAK33100.1; JOINED.
DR EMBL; AY029500; AAK33100.1; JOINED.
DR EMBL; AY029501; AAK33100.1; JOINED.
DR EMBL; AY029502; AAK33100.1; JOINED.
DR EMBL; AY029503; AAK33100.1; JOINED.
DR EMBL; BC052251; AAK52251.1; -
DR EMBL; AB011138; BAA25492.1; -
DR Genew; HGNC:13542; ATP10A.
DR MIM; 605855; -
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:0004012; F:phospholipid-translocating ATPase activity; NAS.
DR GO; GO:0008360; P:regulation of cell shape; NAS.
DR InterPro; IPR001757; ATPase_E1-E2.
DR InterPro; IPR005539; Plippase.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00119; CATATPASE.
DR TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
DR TIGRFAMs; TIGR01494; ATPase P-type; 6.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
KW Hydrolase, Transmembrane, Phosphorylation, Magnesium, ATP-binding;
KW Multigene family.
KW DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 87 106 POTENTIAL.
FT DOMAIN 107 110 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 111 128 POTENTIAL.
FT DOMAIN 129 309 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 310 332 POTENTIAL.
FT DOMAIN 333 362 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 363 384 POTENTIAL.
FT DOMAIN 385 1087 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1088 1108 POTENTIAL.
FT DOMAIN 1109 1119 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1120 1140 POTENTIAL.
FT DOMAIN 1141 1170 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1171 1192 POTENTIAL.
FT DOMAIN 1193 1199 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1200 1222 POTENTIAL.
FT DOMAIN 1223 1249 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1250 1267 POTENTIAL.
FT DOMAIN 1268 1292 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1293 1499 POTENTIAL.
FT MOD RES 427 427 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 1031 1031 MAGNESIUM (BY SIMILARITY).
FT METAL 1035 1035 MAGNESIUM (BY SIMILARITY).
FT DOMAIN 467 470 POLY-GLU.
FT CONFLICT 388 388 Q -> R (IN REF. 4).
SQ SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
Query Match 75.0%; Score 36; DB 1; Length 1499;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 EHVVPXGXSVS 11
DB 469 EHVVPXGXSVS 479

RESULT 2
ID CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
DN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OC NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fenslein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: Belongs to the carb family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

EMBL; AB010554; AAL94625.1; ALT_INIT.	
HAWAP; MF_01210; -; 1.	
InterPro; IPR006275; CarA_L_glu.	
InterPro; IPR005483; CPase_L.	
InterPro; IPR005479; CPase_L_D2.	
InterPro; IPR005480; CPase_L_D3.	
InterPro; IPR005481; CPase_L_N.	
InterPro; IPR004362; MGS_like.	
Pfam; PF00283; CPasease_L_chain; 2.	
Pfam; PF02786; CPasease_L_D2; 2.	
Pfam; PF02787; CPasease_L_D3; 1.	
Pfam; PF02142; MGS; 1.	
PRINTS; PR00098; CPASE.	
TIGRFAMS; TIGR01369; CPaseeII_lig; 1.	
PROSITE; PS00866; CPASE1; 2.	
PROSITE; PS00867; CPASE2; 2.	
Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;	
ATP-binding; Manganese; Complete proteome.	
DOMAIN 1 401	CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
DOMAIN 402 546	OLIGOMERIZATION DOMAIN.
DOMAIN 547 929	CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
DOMAIN 930 1058	ALLOSTERIC DOMAIN.
REPEAT 1 546	
REPEAT 547 1058	ATP (POTENTIAL).
NP_BIND 153 210	ATP (POTENTIAL).
NP_BIND 302 352	MANGANESE 1 (BY SIMILARITY).
METAL 284 284	MANGANESE 1 AND 2 (BY SIMILARITY).
METAL 298 298	MANGANESE 2 (BY SIMILARITY).
METAL 300 300	MANGANESE 3 (BY SIMILARITY).
METAL 820 820	MANGANESE 3 (BY SIMILARITY).
METAL 832 832	SEQUENCE 1058 AA: ED7037AF77C1E39F CRC64;

```

Query Match          72.9%; Score 35; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

      Y      Y      2 EVVPXGXSYS 11
      |||  |||  |||
      190 EIVPGLNYS 199

      Y      Y      2 EVVPXGXSYS 11
      |||  |||  |||
      190 EIVPGLNYS 199

RESULT 3
Y1LK TYDVA
ID Y1LK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 25, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
EN
CS Tobacco yellow dwarf virus (strain Australia) (TYDV).
CC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
CX NCBI_TaxID=31599;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=921188538; PubMed=1546459;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RT. Virology 187:633-642(1992).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC altered, and the source is acknowledged.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL; M81103; AAA47947.1; -.
CC	PIR; A42452; A42452.
DR	InterPro; IPR002621; Gemini_mov.
DR	Pfam; PF01708; gemini_mov; 1.
DR	Hypothetical protein.
KW	SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF58B67 CRC64;
SO	

```

Query Match      70.8%; Score 34; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. NO. 2.2;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 EWVPGKXSYS 11
      :|||:|:|
      7 QVPSGINYS 16

RESULT 4
YJ49 ARCFU
ID YJ49 ARCFU STANDARD; PRT; 165 AA.
AC O28330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AFI1949.
GN AFI1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.

```

SEQUENCE FROM N.A.
STRAIN=VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=98049343; PubMed=9389475;
Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Kerthum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Gloeck A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon *Archaeoglobus fulgidus*.;
Nature 390:364-370(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@sib-sib.ch).

CC	EMBL; A5000968; AAB89307.1; -
DR	PIR; D69493; D69493.
DR	TIGR; AFI949; -
DR	TIGR; AFI949; -
KW	Hypothetical protein; Transmembrane; Complete proteome.
FT	POTENTIAL.
FT	TRANSMEM 7 27
FT	TRANSMEM 141 161
FT	TRANSMEM 165 181
SO	SEQUENCE 17588 MW: BCC17054810ABDF8 CRC664;

```

Query Match      70.8%; Score 34; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qv      1 EEWXPXGXSY 10

```



```

Db
-----
RESULT 5
GSR2_HUMAN STANDARD; PRT; 478 AA.
AC Q9NZM5; Q9NRC6; Q9HAX6; Q9NRP1; Q9NRP4; Q9UFI2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Teki K., Billings S., Ramasamy S., Mohrenweiser H.W.,
RA Schethauer B.W., Louis D.N., Jenkins R.B.,
RA "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT region."
RL Genomics 64:44-50(2000).
[2]
SEQUENCE FROM N.A.
TISSUE=Muscle;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollath S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy E., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]
SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
MEDLINE=99214318; PubMed=10196275;
RA Bruni R., Fineschi B., Ogle W.O., Roizman B.;
RT "A novel cellular protein, p60, interacting with both herpes simplex
RT virus 1 regulatory proteins ICP22 and ICP0 is modified in a
RT cell-type-specific manner and is recruited to the nucleus after
RT infection."
J. Virol. 73:3810-3817(1999).
[4]
SEQUENCE OF 12-478 FROM N.A.
RA Andreu N., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE OF 218-477 FROM N.A.
TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed at high levels in heart and
CC pancreas, moderate levels in placenta, liver, skeletal muscle, and
CC kidney, and low levels in brain and lung.

```

```

CC
CC -1- SIMILARITY: Belongs to the GLTSCR2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AF182076; AAF62873.1; -
CC EMBL; BC004229; AAH04229.1; -
CC EMBL; BC006311; AAH06311.1; -
CC EMBL; BC010095; AAH10095.1; -
CC EMBL; AF296124; AAG30413.1; -
CC EMBL; AL359335; CAB94786.1; -
CC EMBL; AL359335; CAB94787.1; -
CC EMBL; AL122063; CAB95242.1; -
CC SWISS-2DPAGE; Q9NZM5; HUMAN.
CC Genew; HGNC:4333; GLTSCR2.
CC MIM; 605691; -
CC GO; GO:0005622; C:intracellular; NAS.
CC Nuclear protein; Polymorphism.
CC VARIANT 389 389
CC R -> Q.
CC GCS -> HEG (IN REF. 2; AAH04229).
CC G -> R (IN REF. 3).
CC RRKQLWEKLAQGLPREVRAQARLLNPSATRAKPGPD
CC TVERP -> SGRSYGRSWFSRASSPGGAPSPVAQPCN
CC KGNPAPGHRIA (IN REF. 3).
CC SDNPLDRPLVGQDEFFLE -> LNNPKPVVPGCLPFG
CC (IN REF. 3).
CC A -> S (IN REF. 2; AAH04229).
CC D -> H (IN REF. 3).
CC PEGNILDRFKSFQRRNMIEPRERAKFKRYKVKLVKRAP
CC REIQ -> VLTVCRCGAPCPVMTSPSLPVPFRGYRHGCP
CC WAGPVGMPRG (IN REF. 5).
CC EGNILDRFKSFQRRNMIEPRERAKFKRYKVKLVKRAFR
CC EIQL -> RGHSPETGSRAPRG (IN REF. 3).
CC
CC QUERY MATCH 70.8%; Score 34; DB 1; Length 478;
CC Best Local Similarity 60.0%; Pred. No. 11;
CC Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 2 EVVPGKXSYS 11
CC DB 239 EVAPAGASYN 248
CC
CC RESULT 6
CC REV_SIVCZ STANDARD; PRT; 124 AA.
CC AC P17280;
CC DT 01-AUG-1990 (Rel. 15, Created)
CC DT 01-AUG-1990 (Rel. 15, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE REV protein (Anti-repression transactivator protein) (ART/TRS).
CC GN REV.
CC OS Chimpanzee immunodeficiency virus (SIV(cp2)) (CIV).
CC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
CC OX NCBI_TaxID=11723;
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE=90259077; PubMed=2188136;
CC RX Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
CC RT "Genetic organization of a chimpanzee lentivirus related to HIV-1."
CC RL Nature 345:356-359(1990).
CC CC -1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
CC -1- PTM: Phosphoprotein whose state of phosphorylation is mediated by
CC a specific serine kinase activity present in the nucleus.
CC

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; X52154; CAA36405.1; -.
 CC PIR; S09988; VKLJ51.
 CC HIV; X52154; REV5CPZ.
 CC InterPro; IPR000625; REV_protein.
 CC Pfam; PF00424; REV; 1.
 CC Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 CC SEQUENCE 124 AA; 13701 MW; 15877D1BDP65A7B2 CRC64;

 Query Match 68.8%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 4.6;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

 2Y 2 EVVPGXKSYS 11
 107 ETVPGNGYS 116

 RESULT 7
 TOLB HAEMIN STANDARD; PRT; 427 AA.
 AC P44577; P94811;
 DT 01-NOV-1995 (Rel. 32, Created)
 JT 01-NOV-1995 (Rel. 32, Last sequence update)
 JT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE TOLB protein precursor.
 DE TOLB OR HI0302.
 3N TOLB OR HI0302.
 2S Haemophilus influenzae.
 2C Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 2C Pasteurellaceae; Haemophilus.
 2X NCBI_taxid=727;
 2X [1]
 3P SEQUENCE FROM N.A.
 3C STRAIN=rd / KW20 / ATCC 51907;
 3X MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.P., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
 RL Science 269:496-512 (1995).
 3P SEQUENCE FROM N.A.
 3C STRAIN=1479;
 3X MEDLINE=97050550; PubMed=8921895;
 RA Sen K., Sikkema D.J., Murphy T.F.;
 RA "Isolation and characterization of the Haemophilus influenzae tolQ, tolR, tolA and tolB genes.";
 RL Gene 178:75-81 (1996).
 CC -!- FUNCTION: Involved in the tonB-independent uptake of proteins (By similarity).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: Belongs to the tolB family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

 entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; U32722; AAC22040.1; -.
 CC EMBL; U34470; AAC4597.1; -.
 CC PIR; F64064; F64064.
 CC HSP; P1935; 1CRZ.
 CC TIGR; HI0382; -.
 CC HAMAP; MF 00671; -.
 CC InterPro; IPR007195; TolB_N.
 CC Pfam; PF04052; TolB_N; 1.
 CC TransPort; Protein transport; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 24 427
 FT VARIANT 6 6
 FT VARIANT 14 14
 FT VARIANT 17 19
 FT VARIANT 21 21
 FT VARIANT 79 79
 FT VARIANT 129 129
 FT VARIANT 160 160
 FT VARIANT 237 237
 FT VARIANT 322 322
 FT VARIANT 326 326
 FT VARIANT 328 328
 SQ SEQUENCE 427 AA; 44967 MW; 0882201AE9254B9 CRC64;

 Query Match 68.8%; Score 33; DB 1; Length 427;
 Best Local Similarity 60.0%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

 QY 2 EVVPGXKSYS 11
 103 QVWPGNGYS 112

 RESULT 8
 AROA_VIBCH STANDARD; PRT; 426 AA.
 AC Q9KRB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-phosphoenolpyruvate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR VC1732.
 OS Vibrio cholerae
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 CX NCBI_taxid=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RC MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickley E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tetelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Seilers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";
 RL Nature 406:477-483 (2000).
 CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate = phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway; sixth step.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Belongs to the EPSP synthase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB004251; RA994882.1; -;
CC TIGR; V01732; -;
CC HAMAP; MF_00210; -; 1.
CC InterPro; IPR006264; AroA.
CC InterPro; IPR001986; EPSP synth.
CC Pfam; PF00275; EPSP synthase; 1.
CC ProDom; PD001867; EPSP synthase; 1.
CC TIGRfams; TIGR01356; aroA; 1.
CC PROSITE; PS00104; EPSP SYNTHASE 1; 1.
CC PROSITE; PS00885; EPSP SYNTHASE 2; 1.
CC Aromatic amino acid biosynthesis; Transferase; Complete proteome.
KW Amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 426 AA; 46101 MW; 38852D6483BFE1C3 CRC64;

Query Match 66.7%; Score 32; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXSY 10
DB 223 EFVIPAGQSY 232

RESULT 9
ID BCNS CLOPE STANDARD; PRT; 890 AA.
AC P08696;
DT 01-JAN-1988 (Rel. 06, Created)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
DE binding protein 1) (HIV-BP1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE [PROTID-BP1].
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CPN50;
RX MEDLINE=88336297; PubMed=2901768;
RA Garnier T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garnier T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium
RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene.";
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RA STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro.";
RL Mol. Microbiol. 2:607-614(1988).
CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -1- INDUCTION: By UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14481; AAA98248.1; -;
CC EMBL; M32882; AAA98249.1; -;
CC PIR; A30481; A30481.
CC InterPro; IPR000834; Peptidase_M14.
CC InterPro; IPR003646; SH3_Bac.
CC Pfam; PF00246; Zn_carbOpept; 1.
CC SMART; SM00287; SH3b; 3.
CC Antibiatic; Bacteriocin; Plasmid.
KW DOMAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;

Query Match 66.7%; Score 32; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSY 10
DB 170 EVVPGGFTY 178

RESULT 10
ID ZEP1 HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
DE binding protein 1) (HIV-BP1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE [PROTID-BP1].
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90169514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger
RT motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
RA Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger
RT from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142.
RX MEDLINE=92232684; PubMed=1567844;
RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGATTTC-3', WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -----

CC -|- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-
CC glucose + H₂O.
CC
CC -|- COFACTOR: NAD.
CC
CC -|- PATHWAY: DTDLP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC
CC -|- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. DTDLP-GLUCOSE
CC DEHYDRATASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56793; CMA40115.1; -.
CC EMBL; AE008792; AAL21001.1; -.
CC PIR; S15299; S15299.
CC PDB; 1G1A; 21-MAR-01.
CC PDB; 1KEU; 25-JAN-02.
CC PDB; 1KEW; 25-JAN-02.
CC StyGene; SG10345; rfbB.
CC InterPro; IPR005888; dTDP-gluc.dehyt.
CC InterPro; IPR001509; Epimerase_Dh.
CC Pfam; PF01370; Epimerase; 1.
CC TIGRFAMs; TIGR01181; dTDP gluc dehyt; 1.
CC Lipopolysaccharide biosynthesis; Lyase; NAD; Complete proteome;
CC 3D-structure.
CC NP BIND 7 13 NAD (POTENTIAL).
CC SEQUENCE 361 AA; 40718 MW; 3A574B4D917B5C57 CRC64;
Query Match 64.6%; Score 31; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPKXGSX 10
Db :|||
278 DEIVPRATSY 287
RESULT 13
THD1 BURCE
ID THD1 BURCE STANDARD; PRT; 507 AA.
AC P33607;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.3.1.19) (Threonine
DE deaminase).
GN ILVA.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RA Bartell J.B., Lessie T.G.;
RC STRAIN=17616;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
CC threonine in a two-step reaction. The first step is a dehydration
CC of threonine, followed by rehydration and liberation of ammonia.
CC -|- CATALYTIC ACTIVITY: L-threonine = 2-oxobutanoate + NH₃.
CC
CC -|- COFACTOR: Pyridoxal phosphate.
CC
CC -|- PATHWAY: Isoleucine biosynthesis; first step.
CC
CC -|- SUBUNIT: Homotrimer (By similarity).
CC
CC -|- SIMILARITY: Belongs to the serine/threonine dehydratase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40630; AAA83215.1; -.
CC HSSP; P04968; 1TDJ.
CC InterPro; IPR001926; B6 enzyme beta.
CC InterPro; IPR006334; S/T dehydratase BS.
CC InterPro; IPR005787; Thr_dehydratase.
CC InterPro; IPR001721; ThrDh_C.
CC Pfam; PF00291; PALP; 1.
CC Pfam; PF00585; Thr_dehydrat_C; 2.
CC TIGRFAMs; TIGR01124; ilvA_2cterm; 1.
CC PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
CC Isoleucine biosynthesis; Lyase; Pyridoxal phosphate.
CC BINDING 52 52 PYRIDOXAL PHOSPHATE.
CC SEQUENCE 507 AA; 55326 MW; E9A5D110B0597664 CRC64;
Query Match 64.6%; Score 31; DB 1; Length 507;
Best Local Similarity 60.0%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVPKXGSX 11
Db :|||
120 EVIQGESYS 129
RESULT 14
SYL LACLA
ID SYL LACLA STANDARD; PRT; 829 AA.
AC Q9CHB6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR LL0816.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403".
RL Genome Res. 11:731-753(2001).
CC -|- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA (Leu) = AMP +
CC diphosphate + L-leucyl-tRNA (Leu).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC
CC -|- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AS006315; AAK04914.1; -.
CC PIR; H86726; H86726.
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-trNA-syntla.
CC InterPro; IPR002300; tRNA-synt 1a.
CC InterPro; IPR001412; tRNA-synt 1.
CC InterPro; IPR009008; VALRS_ILeRS_edit.
CC Pfam; PF00133; tRNA-synt 1_1.
CC PRINTS; PRO0985; TRNASYNTHLEU.
CC TIGRFAMs; TIGR00396; leus_pact; 1.
CC PROSITE; PS00178; AA tRNA_LIGASE_I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW

```

DR InterPro; IPR002302; Leu-TRNAsyntla.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR009008; ValRS_1fRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
DR KX Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
DR SITE 41 52 "HIGH" REGION.
FT SITE 610 614 "RMSKS" REGION.
FT BINDING 613 613 ATP (BY SIMILARITY).
SQ SEQUENCE 833 AA; 2C6F281AE3D3A896 CRC64;
Query Match 64.6%; Score 31; DB 1; Length 833;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps
QY 1 EEVVPXGXS 9
DB 169 EEVLPDGTs 177
| | | | |
| | | | |
| | | | |

Search completed: June 3, 2004, 11:49:55
Job time : 5.85667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds

(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 EYVVPXGXSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_nhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rv.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	77.1	471	11	Q8R126
2	37	77.1	484	11	Q8VD18
3	37	77.1	484	11	Q8ETX4
4	37	77.1	484	11	Q8BK35
5	37	77.1	1044	16	Q8DIH0
6	36	75.0	3472	1	Q74056
7	35	72.9	225	10	Q40129
8	35	72.9	344	16	Q815A7
9	34	70.8	143	17	Q8TX62
10	34	70.8	174	10	Q9W3T4
11	34	70.8	479	4	Q96CS0
12	34	70.8	541	16	Q98BP5
13	34	70.8	678	12	Q9EIX6
14	34	70.8	1028	16	Q8YJ11
15	34	70.8	1123	16	Q8EWD4
16	34	70.8	1442	17	Q96YH5

17	33	68.8	78	6	Q9XST4
18	33	68.8	175	6	Q81033
19	33	68.8	200	16	Q8XHX3
20	33	68.8	215	6	Q81031
21	33	68.8	217	4	Q00404
22	33	68.8	281	6	Q867A5
23	33	68.8	297	6	Q8HXY9
24	33	68.8	299	4	Q9UEE9
25	33	68.8	415	16	Q7VGY0
26	33	68.8	424	13	Q7ZTH8
27	33	68.8	492	5	Q7YXK8
28	33	68.8	495	11	Q8CID7
29	33	68.8	526	5	Q9VMN9
30	33	68.8	815	10	Q9SF93
31	33	68.8	815	10	Q81850
32	33	68.8	899	16	Q8G4I5
33	33	68.8	933	5	Q8SS39
34	33	68.8	1327	16	Q8EFM1
35	33	68.8	1399	16	Q889X7
36	32	66.7	96	3	Q9X855
37	32	66.7	219	17	Q971S2
38	32	66.7	227	16	Q9RZU8
39	32	66.7	245	16	Q7V6Q4
40	32	66.7	273	16	Q81JQ6
41	32	66.7	273	16	Q814N0
42	32	66.7	290	4	Q86WU1
43	32	66.7	330	16	Q98D00
44	32	66.7	387	16	Q98FX1
45	32	66.7	387	16	Q92MD6

ALIGNMENTS

RESULT 1

Q8R126 PRELIMINARY; PRT; 471 AA.
 ID Q8R126
 AC Q8R126;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN GLTSCR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC025810; AAB25810.1; -
 DR MGD; MGI:2154441; Glta2r2.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 77.1%; Score 37; DB 11; Length 471;
 Best Local Similarity 60.0%; Pred. No. 14;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EYVVPXGXSYS 11
 ||:|:|:|:
 Db 226 EVTPAGASYN 235

RESULT 2

Q8VD18 PRELIMINARY; PRT; 484 AA.
 ID Q8VD18
 AC Q8VD18;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AW536441.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Salivary gland;
RC Strausberg R.;
RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RL ENBL; BC017637; AAH17637.1; --
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55835 MW; BB45F3B4BE02A36 CRC64;

Query Match 77.1%; Score 37; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSYS 11
||:|:|:|:
Db 239 EVIPAGASYN 248

RESULT 3
Q8BTX4 PRELIMINARY; PRT; 484 AA.
ID Q8BTX4
AC Q8BTX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NOD; TISSUE=Thymus;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR ENBL; AK088461; BAC40367.1; --
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EECAD8 CRC64;

Query Match 77.1%; Score 37; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSYS 11
||:|:|:|:
Db 239 EVIPAGASYN 248

RESULT 4
Q8BK35 PRELIMINARY; PRT; 484 AA.
ID Q8BK35
AC Q8BK35;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Pituitary;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR ENBL; AK077341; BAC36760.1; --
DR MGD; MGI:2154441; Gltscr2.
SQ SEQUENCE 484 AA; 55792 MW; BB67949BCBB92D44 CRC64;

Query Match 77.1%; Score 37; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSYS 11
||:|:|:|:
Db 239 EVIPAGASYN 248

RESULT 5
Q8DIH0 PRELIMINARY; PRT; 1044 AA.
ID Q8DIH0
AC Q8DIH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
DE TLL1618.
GN Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BP-1;
RC MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1";
RL DNA Res. 9:123-130(2002).
DR ENBL; AP005374; BAC09170.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001036; Acflvin_res.
DR InterPro; IPR004764; HAEI.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

Query Match 77.1%; Score 37; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXSYS 11
||:|:|:|:
Db 843 EVLPNGIGYS 853
RESULT 6
O74056 PRELIMINARY; PRT; 3472 AA.
ID O74056
AC O74056;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 Hypothetical protein.
 Cenarchaeum symbiosum.
 Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 Cenarchaeum.
 NCBI_TaxID=46770;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=B.
 MEDLINE=98422450; PubMed=9748430;
 Schleper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 Swanson R.V.,
 "Genomic analysis reveals chromosomal variation in natural populations
 of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
 J. Bacteriol. 180:5003-5009 (1998).
 EMBL: AF083072; AAC62699.1; -
 PIR: T31308; T31308.
 GO: GO:0016020; C:membrane; IEA.
 GO: GO:0005215; F:transporter activity; IEA.
 GO: GO:0006810; P:transport; IEA.
 InterPro: IPR000515; BPD transp.
 InterPro: IPR001580; WD40.
 SMART: SMC0320; WD40; 2.
 PROSITE: PS00402; BPD_TRANS_INN_MEMBER; 1.
 Hypothetical protein.
 SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;
 Query Match 75.0%; Score 36; DB 1; Length 3472;
 Best Local Similarity 54.5%; Pred. No. 2; Indels 0; Gaps 0;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 1 EVVVPXGXSYS 11
 2294 EDVIPRGISPS 2304

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EVVVPXGXSYS 11
 DB 32 DEYVPGKTYA 42
 RESULT 8
 Q815A7 PRELIMINARY; PRT; 344 AA.
 Q815A7
 01-JUN-2003 (TrEMBLrel. 24, Created)
 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 ABC transporter substrate-binding protein.
 BC5259.
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=226900;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=22608415; PubMed=12721630;
 Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
 Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
 Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
 Grechkin Y., Pusch G., Haselkorn R., Fongstein M., Ehrlich S.D.,
 Overbeek R., Kyrpides N.;
 "genome sequence of Bacillus cereus and comparative analysis with
 Bacillus anthracis.";
 Nature 423:87-91 (2003).
 EMBL: AE017015; AAP12123.1; -
 InterPro: IPR000437; Prok_lipoprot_S.
 PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 Complete proteome.
 SEQUENCE 344 AA; 38539 MW; C55268ACB7225995 CRC64;
 Query Match 72.9%; Score 35; DB 16; Length 344;
 Best Local Similarity 60.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 1 EVVVPXGXSYS 10
 152 EETAPLGSLY 161
 RESULT 9
 Q8TX62 PRELIMINARY; PRT; 143 AA.
 Q8TX62
 01-JUN-2002 (TrEMBLrel. 21, Created)
 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 Uncharacterized conserved protein.
 MK0814.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 Methanopyrus.
 NCBI_TaxID=2320;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=AV19 / DSM 6324 / JCM 9639;
 MEDLINE=21927647; PubMed=11930014;
 Slesarev A.I., Mezheva V.V., Makarova K.S., Polushin N.N.,
 Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 Natale D.A., Rogozin I.B., Tatusov R.I., Wolf Y.I., Stetter K.O.,
 Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
 "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 and monophyly of archaeal methanogens.";
 Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
 EMBL: AE010372; AA02027.1; -
 InterPro: IPR001602; UPF0047.
 Pfam: PF01894; UPF0047; 1.
 ProDom: PD005232; UPF0047; 1.

DR TIGRFBMS; TIGR00149; TIGR00149; 1.
 DR PROSITE; PS01314; UPF0047; 1.
 DR FIRSP; FIRSF004681; UPF0047; 1.
 KW Complete proteome.
 SQ SEQUENCE 143 AA; 15734 MW; 4C8B28A1FBBEDD0B CRC64;

Query Match 70.8%; Score 34; DB 17; Length 143;
 Best Local Similarity 60.0%; Pred. No. 18;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGXSY 10
 |||||
 Db 75 EELVPGAGY 84

RESULT 10

Q9M3T4 PRELIMINARY; PRT; 174 AA.

AC Q9M3T4;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Acidic endochitinase (EC 3.2.1.14) (Fragment).
 GN FR3A.
 OS Betula verrucosa (White birch) (Betula pendula).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; rosids;
 OC eurosids I; Fagales; Betulaceae; Betula.
 OC NCBI_TaxID=3505;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=leaf;
 RX Hilovara-Tanjo M., Korhonen M.S., Palva T.E., Kangasjarvi J.,
 RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RA EMBL; A2379692; CAB66334.1; --
 DR HSP; P23472; 2HVM.
 DR GO; GO:000843; F:endochitinase activity; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR InterPro; IPR001579; Glyco_hydro_18S.
 DR DR PROSITE; PS01095; CHITINASE_18; 1.
 DR DR PROSITE; PS01095; CHITINASE_18; 1.
 KW Glycosidase; Hydrolase.
 FT NON_TER 174
 FT NON_TER 174
 SQ SEQUENCE 174 AA; 17936 MW; 834ADCC6B5C76634 CRC64;

Query Match 70.8%; Score 34; DB 10; Length 174;
 Best Local Similarity 77.8%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVFXGXSY 11
 |||||
 Db 74 VVFGGGSY 82

RESULT 11

Q96CS0 PRELIMINARY; PRT; 479 AA.

AC Q96CS0; Q96IT7;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Glioma tumor suppressor candidate region gene 2 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung and Skin;
 RA Strausberg R.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014009; AAH14009.1; --
 DR EMBL; BC007248; AAH07248.1; --
 FT NON_TER 1
 SQ SEQUENCE 479 AA; 54529 MW; 0A6C3A8B476F7B8F CRC64;

Query Match 70.8%; Score 34; DB 4; Length 479;
 Best Local Similarity 60.0%; Pred. No. 67;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGXSY 11
 |||||
 Db 240 EVAFAGASY 249

RESULT 12

Q98BP5 PRELIMINARY; PRT; 541 AA.

AC Q98BP5;
 DT 01-OCT-2001 (T-EMBLrel. 18, Created)
 DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Probable DNA ligase.
 GN ML5481.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OC NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003006; BAB51927.1; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0016874; P:ligase activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006280; P:DNA replication; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR Pfam; PF01068; DNA_ligase; 1.
 DR Pfam; PF04679; DNA_ligase_A_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 DR PROSITE; PS01160; DNA_LIGASE_A3; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 541 AA; 60645 MW; 2EFEF705453F28F8 CRC64;

Query Match 70.8%; Score 34; DB 16; Length 541;
 Best Local Similarity 60.0%; Pred. No. 76;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSY 10
 |||||
 Db 445 BELVPGKAY 454

RESULT 13

Q9E1X6 PRELIMINARY; PRT; 678 AA.

AC Q9E1X6;
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Cercopithecine herpesvirus 7.

CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC Alphaherpesvirinae; Varicellovirus.

2X NCBI_TaxID=35245;
2N [1]
RP SEQUENCE FROM N.A.
RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
RT "Complete sequence of the Simian Varicella Virus Genome.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275348; AAG27217.1; -.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0006323; P:DNA packaging; IEA.
DR InterPro; IPR007640; Herpes_UL17.
DR Pfam; PF04559; Herpes_UL17; 1.
KW Hypothetical protein.
SQ SEQUENCE 678 AA; 75850 MW; A17B09E30512FE3C CRC64;

Query Match 70.8%; Score 34; DB 12; Length 678;
Best Local Similarity 50.0%; Pred. No. 98;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

2Y 1 EVVVPXGXS 10
||:|:|
2b 147 EEIIPKGTGY 156

RESULT 14
281J11 PRELIMINARY; PRT; 1028 AA.
ID Q8YU11
AC Q8YU11;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ATP-dependent DNA helicase.
GN BME102/5.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kypides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009470; AAL51457.1; -.
DR PIR; AF3286; AF3286.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent Helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldenhyde_dehydr.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 70.8%; Score 34; DB 16; Length 1028;
Best Local Similarity 54.5%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGXS 11
|:|:|
Db 76 EKIVPPGARYS 86

RESULT 15
Q8EWD4 PRELIMINARY; PRT; 1123 AA.
ID Q8EWD4
AC Q8EWD4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MYB 2560 paralog, 57%.
GN MYB2710.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004171; BAC44062.1; -.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR007326; Lipoprotein_17.
DR Pfam; PF04200; Lipoprotein_17; 3.
KW Complete proteome.
SQ SEQUENCE 1123 AA; 123636 MW; A4D70730E3DB4AC CRC64;
Query Match 70.8%; Score 34; DB 16; Length 1123;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGXS 11
|:|:|
Db 658 EYVPMGLSYS 667

Search completed: June 3, 2004, 11:57:34
Job time : 30.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

title: US-09-909-164-44

effect score: 52

sequence: 1 ZEVVPXGHS 11

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	94.2	11	5	ABB80546 Hepatitis
2	49	94.2	11	5	ABB80534 Hepatitis
3	49	94.2	11	5	ABB80550 Hepatitis
4	49	94.2	11	5	ABB80555 Hepatitis
5	48	92.3	11	5	ABB80523 Hepatitis
6	48	92.3	11	5	ABB80558 Hepatitis
7	48	92.3	11	5	ABB80537 Hepatitis
8	48	92.3	11	5	ABB80560 Hepatitis
9	48	92.3	11	5	ABB80527 Hepatitis
10	48	92.3	11	5	ABB80541 Hepatitis
11	48	92.3	11	5	ABB80532 Hepatitis
12	48	92.3	11	5	ABB80531 Hepatitis
13	40	76.9	11	5	ABB80548 Hepatitis
14	40	76.9	11	5	ABB80549 Hepatitis
15	40	76.9	11	5	ABB80547 Hepatitis
16	40	76.9	11	5	ABB80544 Hepatitis
17	40	76.9	11	5	ABB80556 Hepatitis
18	40	76.9	11	5	ABB80557 Hepatitis
19	40	76.9	11	5	ABB80551 Hepatitis
20	40	76.9	11	5	ABB80553 Hepatitis
21	40	76.9	11	5	ABB80552 Hepatitis
22	40	76.9	11	5	ABB80545 Hepatitis
23	39	75.0	11	5	ABB80525 Hepatitis
24	39	75.0	11	5	ABB80534 Hepatitis
25	39	75.0	11	5	ABB80561 Hepatitis

26	39	75.0	11	5	ABB80521 Hepatitis
27	39	75.0	11	5	ABB80522 Hepatitis
28	39	75.0	11	5	ABB80536 Hepatitis
29	39	75.0	11	5	ABB80566 Hepatitis
30	39	75.0	11	5	ABB80542 Hepatitis
31	39	75.0	11	5	ABB80543 Hepatitis
32	39	75.0	11	5	ABB80563 Hepatitis
33	39	75.0	11	5	ABB80565 Hepatitis
34	39	75.0	11	5	ABB80524 Hepatitis
35	39	75.0	11	5	ABB80533 Hepatitis
36	39	75.0	11	5	ABB80529 Hepatitis
37	39	75.0	11	5	ABB80535 Hepatitis
38	39	75.0	11	5	ABB80567 Hepatitis
39	39	75.0	11	5	ABB80528 Hepatitis
40	39	75.0	11	5	ABB80538 Hepatitis
41	39	75.0	11	5	ABB80540 Hepatitis
42	39	75.0	11	5	ABB80562 Hepatitis
43	39	75.0	11	5	ABB80559 Hepatitis
44	39	75.0	11	5	ABB80526 Hepatitis
45	39	75.0	11	5	ABB80530 Hepatitis

ALIGNMENTS

RESULT 1

ABB80546

ID ABB80546 standard; peptide; 11 AA.

XX ABB80546;

XX

XX 08-OCT-2002 (first entry)

DT

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.

DE

DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX virucide.

XX Synthetic.

XX

XX Key Location/Qualifiers

PH Modified-site 1 /note= "N-terminal acetyl"

PT

PT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

PT

PT Modified-site 11 /note= "C-terminal amide"

PT

XX WO200208251-A2.

XX

XX 31-JAN-2002.

XX

XX 19-JUL-2001; 2001WO-US023169.

XX

XX 21-JUL-2000; 2000US-0220101P.

XX

XX (CORV-) CORVAS INT INC.

XX

XX Lim-Wilby M, Levy OE, Brunck TK;

XX

XX WPI; 2002-361643/39.

XX

XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus

PT

PT

XX Claim 17; Page 65; 69pp; English.

XX

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 94.2%; Score 49; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0045; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 2
 ABB80554
 ID ABB80554 standard; peptide; 11 AA.

AC ABB80554;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT

WO200208251-A2.
 31-JAN-2002.

19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.
 Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.
 Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

Query Match 94.2%; Score 49; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0045;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 3
 ABB80550
 ID ABB80550 standard; peptide; 11 AA.

AC ABB80550;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT

WO200208251-A2.
 31-JAN-2002.

19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.
 Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;

Query Match 94.2%; Score 49; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0045;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 4

XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.
DE	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX	virucide.
KW	Synthetic.
OS	
XX	Key Location/Qualifiers
FH	Modified-site 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 9 /note= "D-form residue"
FT	Modified-site 11 /note= "C-terminal amide"
PN	WO200208251-A2.
XX	31-JAN-2002.
PD	19-JUL-2001; 2001WO-US023169.
XX	21-JUL-2000; 2000US-0220101P.
PR	(CORV-) CORVAS INT INC.
XX	Lim-Wilby M, Levy OE, Brunck TK;
PI	WPI; 2002-361643/39.
DR	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
PT	Claim 17; Page 64; 69pp; English.
PS	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
CC	Sequence 11 AA;
SQ	Query Match 92.3%; Score 48; DB 5; Length 11; Best Local Similarity 90.9%; Pred. No. 0.0072; Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0
OY	1 EEVVPXGXHS 11
DB	1 EEVVPXGXHS 11
RESULT 6	
ABB80558	ID ABB80558 standard; peptide; 11 AA.
XX	AC ABB80558;
XX	DT 08-OCT-2002 (first entry)
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW	virucide.
XX	Synthetic.
OS	
XX	Key Location/Qualifiers
FH	Modified-site 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 8 /note= "D-form residue"
FT	Misc-difference 9 /note= "D-form residue"
FT	Modified-site 11 /note= "C-terminal amide"
PN	WO200208251-A2.
XX	31-JAN-2002.
PD	19-JUL-2001; 2001WO-US023169.
XX	21-JUL-2000; 2000US-0220101P.
PR	(CORV-) CORVAS INT INC.
XX	Lim-Wilby M, Levy OE, Brunck TK;
PI	WPI; 2002-361643/39.
DR	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
PT	Claim 17; Page 65; 69pp; English.
PS	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
CC	Sequence 11 AA;
SQ	Query Match 94.2%; Score 49; DB 5; Length 11; Best Local Similarity 90.9%; Pred. No. 0.0045; Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DY	1 EEVVPXGXHS 11
DB	1 EEVVPXGXHS 11
RESULT 5	
ABB80523	ID ABB80523 standard; peptide; 11 AA.
XX	AC ABB80523;
XX	DT 08-OCT-2002 (first entry)

```

FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX PN
XX XX
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PA
XX PI
XX DR
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PA
XX PI
XX DR
XX WPI; 2002-361643/39.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX XX
XX SQ Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.0072;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
Db ||||| |||
1 EEVVPXGMHYS 11

RESULT 7
ABB80537
ID ABB80537 standard; peptide; 11 AA.
XX AC
XX ABB80537;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.

```

```

FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
XX PN
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PA
XX PI
XX DR
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX PS
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX XX
XX SQ Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.0072;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHYS 11
Db ||||| |||
1 EEVVPXGMHYS 11

RESULT 8
ABB80560
ID ABB80560 standard; peptide; 11 AA.
XX AC
XX ABB80560;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.

```

XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.0072;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EEVVPXGXHYS 11
XX |||||
XX 1 EEVVPXGXHYS 11
XX
XX RESULT 9
XX ABB80527
XX ID ABB80527 standard; peptide; 11 AA.
XX AC ABB80527;
XX DT 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 5 /note= "N-terminal acetyl"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 5; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.0072;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EEVVPXGXHYS 11
XX |||||
XX 1 EEVVPXGXHYS 11
XX
XX RESULT 10
XX ABB80541
XX ID ABB80541 standard; peptide; 11 AA.
XX AC ABB80541;
XX DT 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

CC	pharmaceutical composition comprising the peptide as an active ingredient	92.3%;	Score 48;	DB 5;	Length 11;	
XX	is useful for treating disorders associated with hepatitis C virus	Best Local Similarity 90.9%;	Pred. No. 0.0072;			
XX		Matches 10;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	Sequence 11 AA;	1 EEVVPXGXHYS 11				
Db		1 EEVVPXGGHYS 11				
RESULT 12						
AB880531						
ID	AB880531 standard; peptide; 11 AA.					
XX	AC	AB880531;				
XX	DT	08-OCT-2002 (first entry)				
XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.					
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;					
XX	virucide.					
OS	Synthetic.					
XX	Key	Location/Qualifiers				
PH	Modified-site 1	/note= "N-terminal acetyl"				
FT	Modified-site 6					
FT	Modified-site	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"				
FT	Modified-site 11					
FT	Modified-site	/note= "C-terminal amide"				
XX	WC200208251-A2.					
XX	31-JAN-2002.					
XX	19-JUL-2001; 2001WO-US023169.					
XX	21-JUL-2000; 2000US-0220101P.					
XX	(CORV-) CORVAS INT INC.					
XX	Lim-Wilby M, Levy OE, Brunck TK;					
XX	WPI; 2002-361643/39.					
XX	Novel peptide compound having hepatitis C virus protease inhibitory					
PT	activity useful for treating disorders associated with hepatitis C virus					
PT	protease.					
XX	Claim 17; Page 64; 69pp; English.					
XX	The sequence represents a peptide compound of the invention having					
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the					
CC	invention are alpha-ketoamide peptide analogues. The peptides have					
CC	virucide activity, and are useful for treating and in the manufacture of					
CC	a medicament to treat disorders associated with HCV protease. A					
CC	pharmaceutical composition comprising the peptide as an active ingredient					
CC	is useful for treating disorders associated with hepatitis C virus					
XX	Sequence 11 AA;					
QY	Sequence 11 AA;	92.3%;	Score 48;	DB 5;	Length 11;	
Best Match	Best Local Similarity 90.9%;	Pred. No. 0.0072;				
Matches 10;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		

db ||||| |||||
1 EEVVPXGHHYS 11

RESULT 13
ABB80548
ID ABB80548 standard; peptide; 11 AA.
X AC ABB80548;
X DT 08-OCT-2002 (first entry)
X DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
X KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
X OS virucide.
X SS Synthetic.
X PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
X WO200208251-A2.
X 31-JAN-2002.
X 19-JUL-2001; 2001WO-US023169.
X 21-JUL-2000; 2000US-0220101P.
X (CORV-) CORVAS INT INC.
X Lim-Wilby M, Levy OE, Brunck TK;
X WPI; 2002-361643/39.
X Novel peptide compound having hepatitis C virus protease inhibitory
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X Claim 17; Page 65; 69pp; English.
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;
Query Match 76.9%; Score 40; DB 5; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EEVVPXGHHYS 11
|||||
Db 1 EEVVPXGHHYS 11

RESULT 15
ABB80547
ID ABB80547 standard; peptide; 11 AA.
X AC ABB80547;
X DT 08-OCT-2002 (first entry)
X DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
X OS virucide.
X SS Synthetic.
X PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
X WO200208251-A2.
X 31-JAN-2002.
X 19-JUL-2001; 2001WO-US023169.
X 21-JUL-2000; 2000US-0220101P.
X (CORV-) CORVAS INT INC.
X Lim-Wilby M, Levy OE, Brunck TK;
X WPI; 2002-361643/39.
X Novel peptide compound having hepatitis C virus protease inhibitory
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X Claim 17; Page 65; 69pp; English.
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus

Sequence 11 AA;
Query Match 76.9%; Score 40; DB 5; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EEVVPXGHHYS 11
|||||
Db 1 EEVVPXGHHYS 11

RESULT 14
ABB80549
ID ABB80549 standard; peptide; 11 AA.
X AC ABB80549;
X DT 08-OCT-2002 (first entry)
X DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
X OS virucide.
X SS Synthetic.
X PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
X WO200208251-A2.
X 31-JAN-2002.
X 19-JUL-2001; 2001WO-US023169.
X 21-JUL-2000; 2000US-0220101P.
X (CORV-) CORVAS INT INC.
X Lim-Wilby M, Levy OE, Brunck TK;
X WPI; 2002-361643/39.
X Novel peptide compound having hepatitis C virus protease inhibitory
X activity useful for treating disorders associated with hepatitis C virus
X protease.
X Claim 17; Page 65; 69pp; English.
X The sequence represents a peptide compound of the invention having
X hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
X invention are alpha-ketoamide peptide analogues. The peptides have
X virucide activity, and are useful for treating and in the manufacture of
X a medicament to treat disorders associated with HCV protease. A
X pharmaceutical composition comprising the peptide as an active ingredient
X is useful for treating disorders associated with hepatitis C virus

```

XX Synthetic.
XX OS
XX FH Key Location/Qualifiers
XX FT Modified-site 1 /note= "N-terminal acetyl"
XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX FT residue 7"
XX FT Modified-site 11 /note= "C-terminal amide"
XX FT
XX WO200208251-A2.
XX PN
XX PD 31-JAN-2002.
XX XX
XX PF 19-JUL-2001; 2001WO-US023169.
XX XX
XX PR 21-JUL-2000; 2000US-0220101P.
XX XX
XX PA (CORV-) CORVAS INT INC.
XX XX
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX XX
XX DR WPI; 2002-361643/39.
XX XX
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.
XX XX
XX PS Claim 17; Page 65; 69pp; English.
XX XX
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ
XX Sequence 11 AA;

Query Match 76.9%; Score 40; DB 5; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
Db 1 EEVVPXGTDYS 11

```

Search completed: June 3, 2004, 11:48:24
Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-44

Effect score: 52
Sequence: 1 EVVVPXGXHY 11

Coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS-COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	856	4	US-09-252-991A-21444
2	36	69.2	1037	4	US-09-134-001C-4794
3	35	67.3	2277	4	US-09-252-991A-26615
4	34	65.4	126	2	US-08-879-995A-3
5	34	65.4	126	3	US-09-215-096-3
6	33	63.5	63	4	US-09-107-532A-6771
7	33	63.5	447	3	US-08-961-083-182
8	33	63.5	447	4	US-09-536-784-182
9	33	63.5	484	4	US-09-468-656A-6
10	33	63.5	600	2	US-08-821-119-19
11	33	63.5	600	2	US-08-821-118-2
12	33	63.5	763	3	US-08-961-083-66
13	33	63.5	763	4	US-09-536-784-66
14	33	63.5	796	3	US-08-961-083-56
15	33	63.5	796	4	US-09-536-784-56
16	33	63.5	819	4	US-09-468-656A-8
17	33	63.5	819	4	US-09-468-656A-10
18	33	63.5	838	4	US-09-468-656A-4
19	32	61.5	253	4	US-09-328-352-6440
20	32	61.5	323	4	US-09-543-681A-7304
21	32	61.5	738	1	US-08-530-010-3
22	32	61.5	738	1	US-08-530-010-5
23	32	61.5	738	1	US-08-530-010-7
24	32	61.5	738	1	US-08-530-010-9
25	32	61.5	738	1	US-08-530-010-11
26	32	61.5	738	2	US-08-484-101B-3
27	32	61.5	738	2	US-08-484-101B-5

28	61.5	738	2	US-08-484-101B-7	Sequence 7, Appli
29	61.5	738	2	US-08-484-101B-9	Sequence 9, Appli
30	61.5	738	2	US-08-484-101B-11	Sequence 11, Appli
31	61.5	738	2	US-08-714-524D-3	Sequence 3, Appli
32	61.5	738	3	US-08-714-524D-5	Sequence 5, Appli
33	61.5	738	3	US-08-714-524D-7	Sequence 7, Appli
34	61.5	738	3	US-08-714-524D-9	Sequence 9, Appli
35	61.5	738	3	US-08-714-524D-11	Sequence 11, Appli
36	61.5	1247	4	US-08-252-991A-32960	Sequence 32960, A
37	59.6	70	4	US-09-134-001C-3950	Sequence 3950, Ap
38	59.6	237	4	US-09-134-001C-4185	Sequence 4185, Ap
39	59.6	252	4	US-09-328-352-7284	Sequence 7284, Ap
40	59.6	388	4	US-08-178-257-15	Sequence 15, Appli
41	59.6	390	4	US-09-489-039A-8065	Sequence 8065, Ap
42	59.6	502	4	US-09-342-647-4	Sequence 4, Appli
43	59.6	622	2	US-08-459-146-2	Sequence 2, Appli
44	59.6	622	2	US-08-459-065-2	Sequence 2, Appli
45	59.6	630	4	US-09-342-647-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-21444
; Sequence 21444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21444
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 73.1%; Score 38; DB 4; Length 856;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHY 10
Db 64 EAVVPGGHHY 73

RESULT 2
US-09-134-001C-4794
; Sequence 4794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4794
; LENGTH: 1037
; TYPE: PRT

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.2%; Score 36; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
Db 199 KEVWSNGLHY 209

RESULT 3
US-09-252-991A-26615
Sequence 26615, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107195.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26615
LENGTH: 277
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match 67.3%; Score 35; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
Db 48 EETVPGGHTS 58

RESULT 4
US-08-879-995A-3
Sequence 3, Application US/08879995A
Patent No. 5985606
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
US-08-879-995A-3

Query Match 65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXH 9
Db 28 EQVVPGGCH 36

RESULT 5
US-09-215-096-3
Sequence 3, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590

S-09-215-096-3

Query Match 65.4%; Score 34; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Y 1 BEVVPXGXH 9
b 28 EQVPGGHH 36

RESULT 6

S-09-107-532A-6771
Sequence 6771, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6771:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...63
SEQUENCE DESCRIPTION: SEQ ID NO: 6771:
S-09-107-532A-6771

Query Match 63.5%; Score 33; DB 4; Length 63;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHYS 11
b 5 BEINAFGDHYN 15

RESULT 7

US-08-961-083-182

Sequence 182, Application US/08961083

Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 63.5%; Score 33; DB 3; Length 447;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VDPXGXHY 10
b 178 IVPHGHHY 185

RESULT 8

US-09-536-784-182
Sequence 182, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match 63.5%; Score 33; DB 4; Length 447;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
Db 178 IVPHGHHY 185

RESULT 9

US-09-468-656A-6
Sequence 6, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
FILE REFERENCE: 469201-444
CURRENT FILING DATE: 1999-12-02
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6

Query Match 63.5%; Score 33; DB 4; Length 484;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
Db 202 IVPHGHHY 209

RESULT 10

US-08-821-119-19
Sequence 19, Application US/08821119
Patent No. 5821104
GENERAL INFORMATION:
APPLICANT: Holm, Kaj Andre
APPLICANT: Rasmussen, Grethe
APPLICANT: Halkier, Torben
APPLICANT: Lembeck, Jan
TITLE OF INVENTION: Tripeptidyl Aminopeptidase
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58211040 No. 5821104disk of No. 5821104th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,119
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107.204-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-119-19

Query Match 63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGXHY 11
Db 31 VPKGWHYS 38

RESULT 11

US-08-821-118-2
Sequence 2, Application US/08821118
Patent No. 589889
GENERAL INFORMATION:
APPLICANT: Rey, Michael
APPLICANT: Golightly, Elizabeth
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5898890 No. 589889disk of No. 589889th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,118
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4107.400-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-118-2

Query Match 63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

4 VPKGXHY 11
31 VPKGWHYS 38

RESULT 12
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 63.5%; Score 33; DB 3; Length 763;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

3 VVPKXHY 10
184 IVPKGDHY 191

Db 184 IVPKGDHY 191

RESULT 13

US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:

US-09-536-784-66
Query Match 63.5%; Score 33; DB 4; Length 763;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPKXHY 10

Db 184 IVPKGDHY 191

RESULT 14

US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match 63.5%; Score 33; DB 3; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10
Db 185 IVPFGDHY 192

RESULT 15

US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56
Query Match 63.5%; Score 33; DB 4; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 3 VVPXGXHY 10
Db 185 IVPFGDHY 192
Search completed: June 3, 2004, 12:03:09
Job time : 12.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-44
Perfect score: 52
Sequence: 1 EVVFPXGXYHS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/FCR_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/FCRUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	94.2	11	12	US-09-909-164-30
2	49	94.2	11	12	US-09-909-164-34
3	49	94.2	11	12	US-09-909-164-38
4	49	94.2	11	12	US-09-909-164-39
5	48	92.3	11	12	US-09-909-164-7
6	48	92.3	11	12	US-09-909-164-11
7	48	92.3	11	12	US-09-909-164-15
8	48	92.3	11	12	US-09-909-164-16
9	48	92.3	11	12	US-09-909-164-21
10	48	92.3	11	12	US-09-909-164-25
11	48	92.3	11	12	US-09-909-164-42
12	48	92.3	11	12	US-09-909-164-44
13	40	76.9	11	12	US-09-909-164-28
14	40	76.9	11	12	US-09-909-164-29
15	40	76.9	11	12	US-09-909-164-31

16	40	76.9	11	12	US-09-909-164-32	Sequence 32, Appl
17	40	76.9	11	12	US-09-909-164-33	Sequence 33, Appl
18	40	76.9	11	12	US-09-909-164-35	Sequence 35, Appl
19	40	76.9	11	12	US-09-909-164-36	Sequence 36, Appl
20	40	76.9	11	12	US-09-909-164-37	Sequence 37, Appl
21	40	76.9	11	12	US-09-909-164-40	Sequence 40, Appl
22	40	76.9	11	12	US-09-909-164-41	Sequence 41, Appl
23	39	75.0	11	12	US-09-909-164-5	Sequence 5, Appl
24	39	75.0	11	12	US-09-909-164-6	Sequence 6, Appl
25	39	75.0	11	12	US-09-909-164-8	Sequence 8, Appl
26	39	75.0	11	12	US-09-909-164-9	Sequence 9, Appl
27	39	75.0	11	12	US-09-909-164-10	Sequence 10, Appl
28	39	75.0	11	12	US-09-909-164-12	Sequence 12, Appl
29	39	75.0	11	12	US-09-909-164-13	Sequence 13, Appl
30	39	75.0	11	12	US-09-909-164-14	Sequence 14, Appl
31	39	75.0	11	12	US-09-909-164-17	Sequence 17, Appl
32	39	75.0	11	12	US-09-909-164-18	Sequence 18, Appl
33	39	75.0	11	12	US-09-909-164-19	Sequence 19, Appl
34	39	75.0	11	12	US-09-909-164-20	Sequence 20, Appl
35	39	75.0	11	12	US-09-909-164-22	Sequence 22, Appl
36	39	75.0	11	12	US-09-909-164-23	Sequence 23, Appl
37	39	75.0	11	12	US-09-909-164-24	Sequence 24, Appl
38	39	75.0	11	12	US-09-909-164-26	Sequence 26, Appl
39	39	75.0	11	12	US-09-909-164-27	Sequence 27, Appl
40	39	75.0	11	12	US-09-909-164-43	Sequence 43, Appl
41	39	75.0	11	12	US-09-909-164-45	Sequence 45, Appl
42	39	75.0	11	12	US-09-909-164-46	Sequence 46, Appl
43	39	75.0	11	12	US-09-909-164-47	Sequence 47, Appl
44	39	75.0	11	12	US-09-909-164-48	Sequence 48, Appl
45	39	75.0	11	12	US-09-909-164-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1
US-09-909-164-30
; Sequence 30, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-30

Query Match 94.2%; Score 49; DB 12; Length 11;

```

; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)-(8)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-38
Query Match:          94.3%; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGSHYS 11
        |||||
Db       1 EEVVPXGSHYS 11

RESULT 4
US-09-909-164-39
; Sequence 39, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)-(1)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)-(9)

```

OTHER INFORMATION: D-amino acids
JS-09-909-164-39

Query Match 94.28; Score 49; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0036;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXHYS 11
|||||
DB 1 EVVVPXGXHYS 11

RESULT 5

JS-09-909-164-7
Sequence 7, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 7
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXHYS 11
|||||
DB 1 EVVVPXGXHYS 11

RESULT 6

JS-09-909-164-11
Sequence 11, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 11
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
US-09-909-164-11

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXHYS 11
|||||
DB 1 EVVVPXGXHYS 11

RESULT 7

US-09-909-164-15
Sequence 15, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 15
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)

CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patent in version 3.1
SEQ ID NO 15

Query Match 92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXHYS 11
|||||
DB 1 EVVVPXGXHYS 11

RESULT 6

JS-09-909-164-11
Sequence 11, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US

```

; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 21
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-21

Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.006; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1;

QY      1 EEVVPXGXHYS 11
        |||||
Db       1 EEVVPXGXHYS 11

RESULT 10
US-09-909-164-25
; Sequence 25, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 25
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)

```

```

; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
JS-09-909-164-25
Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.006;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

2y 1 BEVVPXGXHYS 11
   |||||
2b 1 BEVVPXGXHYS 11

RESULT 11
JS-09-909-164-42
; Sequence 42, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met(O)
; US-09-909-164-44

Query Match          92.3%; Score 48; DB 12; Length 11;
Best Local Similarity 100.0%; Pred.No. 0.006;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
   |||||
Db 1 BEVVPXGXHYS 11

RESULT 13
US-09-909-164-28
; Sequence 28, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)

2y 1 BEVVPXGXHYS 11
   |||||
2b 1 BEVVPXGXHYS 11

RESULT 12
JS-09-909-164-44
; Sequence 44, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
```

```
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; US-09-909-164-28
```

```
Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EEVVPXGXHYS 11
| | | | | | | |
Db 1 EEVVPXGTSYS 11
```

RESULT 14

```
US-09-909-164-29
; Sequence 29, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-29
```

```
Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EEVVPXGXHYS 11
| | | | | | | |
Db 1 EEVVPXGTSYS 11
```

RESULT 15

```
US-09-909-164-31
; Sequence 31, Application US/09909164
```

```
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; US-09-909-164-31
```

```
Query Match 76.9%; Score 40; DB 12; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EEVVPXGXHYS 11
| | | | | | | |
Db 1 EEVVPXGTSYS 11
```

```
Search completed: June 3, 2004, 12:57:16
Job time : 33.7333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model
un on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

title: US-09-909-164-44
effect score: 52
sequence: 1 EEVVPXGXHY 11

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	35	67.3	233	2 T02590	DNA binding protei
2	35	67.3	308	2 A72207	ftsH proteinase ac
3	35	67.3	360	2 B69086	cell division prot
4	35	67.3	1028	2 AF3286	ATP-dependent DNA
5	35	67.3	1396	2 S36851	L-shaped tail fibe
6	34	65.4	102	2 A42452	V1 protein - tobac
7	34	65.4	126	2 A25905	tachykinin B precu
8	34	65.4	279	2 C75538	hypothetical prote
9	34	65.4	495	2 T28717	hypothetical prote
10	34	65.4	534	2 A69284	coenzyme F420-quin
11	34	65.4	822	2 T46758	hypothetical 92.4k
12	34	65.4	1057	2 F89892	carbamoyl-phosphat
13	33	63.5	46	2 E97985	hypothetical prote
14	33	63.5	124	1 VKL251	trans-regulatory s
15	33	63.5	156	2 D82618	conserved hypothet
16	33	63.5	252	2 AE2001	hypothetical prote
17	33	63.5	412	2 A48702	2-methyl-branched
18	33	63.5	460	2 S69046	hypothetical prote
19	33	63.5	510	2 G86430	T518.1 protein - A
20	33	63.5	743	2 S38143	hypothetical prote
21	33	63.5	802	2 C95116	conserved domain p
22	33	63.5	819	2 B95136	conserved domain p
23	33	63.5	828	2 E98004	hypothetical prote
24	33	63.5	839	2 G95115	conserved hypothet
25	33	63.5	853	2 C97985	hypothetical prote
26	33	63.5	855	2 D98004	histidine Motif-Co
27	33	63.5	1039	2 H95115	conserved hypothet
28	33	63.5	1039	2 D97985	hypothetical prote
29	32	61.5	162	2 A70939	probable monooxygen

30 32 61.5 225 2 SS7810 hypothetical prote
31 32 286 2 C86169 prohibitin 2 [impo
32 32 61.5 311 2 H69194 GMP synthetase, su
33 32 61.5 322 2 AF3211 hypothetical prote
34 32 61.5 364 2 S47540 fructose-bisphosph
35 32 61.5 369 2 T17267 hypothetical prote
36 32 61.5 374 2 G69119 probable aspartate
37 32 61.5 425 2 T24111 hypothetical prote
38 32 61.5 454 2 G96568 probable non-photo
39 32 61.5 519 2 G84598 probable bZIP tran
40 32 61.5 536 2 A13544 aldehyde dehydroge
41 32 61.5 627 2 A96663 DNA mismatch repai
42 32 61.5 696 2 A91247 phage transposase
43 32 61.5 701 2 S61239 hypothetical prote
44 32 61.5 716 1 JC5061 macrophage-stimula
45 32 61.5 738 2 A48246 ethylene-response

ALIGNMENTS

RESULT 1

T02590
DNA binding protein EREBP-2 - common tobacco
C/Species: Nicotiana tabacum (common tobacco)
C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
C/Accession: T02590
R/Ohme-Takagi, M.; Shimshi, H.
Plant Cell 7, 173-182, 1995
A/Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response
A/Reference number: Z14671; MUID:95276459; PMID:7756828
A/Accession: T02590
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-233 <OHX>
A/Cross-references: EMBL:P38126; NID:G790362; PIDD:BA07324.1; PID:G1208498
A/Experimental source: strain BY4; tissue-type leaf

Query Match 67.3%; Score 35; DB 2; Length 233;
Best Local Similarity 60.0%; Pred. No. 9.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 1 EEVVPXGXHY 10
Db 90 QAVVPKGRHY 99

RESULT 2

A72207
ftsH proteinase activity modulator HfLK - Thermotoga maritima (strain MSB8)
C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C/Accession: A72207
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.,
C.M.
Nature 399, 323-329, 1999
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:99287316; PMID:10360571
A/Accession: A72207
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-308 <ARN>
A/Cross-references: GB:AE001819; GB:AE000512; NID:G4982396; PIDD:AAJ36885.1; PID:G498240
A/Experimental source: strain MSB8
C/Genetics:
A/Gene: TMI822
C/Superfamily: erythrocyte band 7 integral membrane protein

Query Match 67.3%; Score 35; DB 2; Length 308;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```

QY      3 VVPXGXHY 10
Db      41 VVPSGIHY 48

RESULT 3
E69086
Cell division protein - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: E69086
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierbowski, J.; Gibson, R.; Jiwani, N.;
; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: E69086
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360 <MTH>
A:Cross-references: GB:AE0000923; GB:AE0000666; NID:G2622766; PIDN:AAB86115.1; PID:G262277
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1642
C:Superfamily: cell division protein M0174

Query Match      67.3%; Score 35; DB 2; Length 360;
Best Local Similarity 45.5%; Pred. No. 16;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEWVPGXHY 11
Db      98 EDLVPMGSHHT 108

RESULT 4
AF3286
ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3286
R:Belvecchio, V.G.; Kapratral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1028 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51457.1; PID:G17982167; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0275
A:Map position: 1

Query Match      67.3%; Score 35; DB 2; Length 1028;
Best Local Similarity 54.5%; Pred. No. 49;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEWVPGXHY 11
Db      76 EKIVPFGARYS 86

RESULT 5
S36851
L-shaped tail fiber protein - phage T5
N:Alternate names: ltf protein
C:Species: phage T5
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C:Accession: S65934; S01984; S36851

R:Kallman, A.V.; Kulshin, V.E.; Shlyapnikov, M.G.; Keenzenko, V.N.; Kryukov, V.M.
FEBS Lett. 366, 46-48, 1995
A:Title: The nucleotide sequence of the bacteriophage T5 ltf gene.
A:Reference number: S65934; MUID:95309401; PMID:7789514
A:Accession: S65934
A:Molecule type: DNA
A:Residues: 1-1396 <KAL>
A:Cross-references: EMBL:X69460; NID:G15415; PIDN:CAA49220.1; PID:G15416
R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
Nucleic Acids Res. 16, 6230, 1988
A:Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early an
A:Reference number: S01982; MUID:88289370; PMID:3267228
A:Accession: S01984
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 934-985, A', 987-1396 <KA2>
A:Cross-references: EMBL:X07559
C:Genetics:
A:Gene: ltf
C:Keywords: late protein; tail fiber

Query Match      67.3%; Score 35; DB 2; Length 1396;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 EEWVPGXHY 11
Db      1360 KTIIPAGDHY 1369

RESULT 6
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yell
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match      65.4%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 EEWVPGXHY 11
Db      7 QVVPFGINYS 16

RESULT 7
A25905
tachykinin B precursor - bovine
N:Alternate names: neuromedin K
C:Species: Bos primigenius taurus (cattle)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25905
R:Kotani, H.; Hoshimaru, M.; Nawa, H.; Nakanishi, S.
Proc. Natl. Acad. Sci. U.S.A. 85, 7074-7078, 1986
A:Title: Structure and gene organization of bovine neuromedin K precursor.
A:Reference number: A25905; MUID:86313713; PMID:3462746
A:Accession: A25905
A:Molecule type: DNA
A:Residues: 1-126 <KOT>
A:Cross-references: GB:M14351; NID:G163587; PIDN:AAA30723.1; PID:G163590
C:Superfamily: neurokinin B precursor
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-126/Product: protachykinin B #status predicted <MAT>

```

```

Query Match      65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 8.2;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Y 1 EVVVPXGXH 9
  |||||
b 28 EQVVPGGGH 36

RESULT 8
75538
;Protein: hypothetical protein - Deinococcus radiodurans (strain R1)
;Species: Deinococcus radiodurans
;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
;Accession: C75538
;Mol. type: M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
; Smith, H.O.; Venter, J.C.; Fraser, C.M.
;Citation: Science 286, 1571-1577, 1999
;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
;Reference number: A75250; MUID:20036996; PMID:10567266
;Accession: C75538
;Status: preliminary
;Molecule type: DNA
;Residues: 1-279 <WHI>
;Cross-references: GB:AE001889; GB:AE000513; NID:g6457944; PIDN:AAF09867.1; PID:g645795
;Experimental source: strain R1
;Genetics:
;Gene: DR0271
;Map position: 1
;Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match      65.4%; Score 34; DB 2; Length 279;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 4 VFXGXHYS 11
  |||||
b 100 VPLGRHYS 107

RESULT 9
28717
;Protein: hypothetical protein F10D2.3 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
;Accession: T28717
;Graves, T.; Wohlmann, P.; Gillam, B.
;Submitted to the EMBL Data Library, September 1997
;Description: The sequence of C. elegans cosmid F10D2.
;Reference number: Z20515
;Accession: T28717
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-495 <GRA>
;Cross-references: EMBL:AF02972; PIDN:AA048234.1; GSPDB:GN00023; CBSP:F10D2.3
;Experimental source: strain Bristol N2; clone F10D2
;Genetics:
;Gene: CBSP:F10D2.3
;Map position: 5
;Introns: 37/2; 90/1; 113/1; 183/3; 356/3; 380/3; 428/2

Query Match      65.4%; Score 34; DB 2; Length 495;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Y 1 BEVVPXGXHY 10
  |||||
b 218 ENIVTQKH 227

RESULT 10
169294

```

```

coenzyme F420-quinone oxidoreductase (EC 1.6.5.-) 56K chain - Archaeoglobus fulgidus
N;Alternate names: sarcosine oxidase alpha chain soxA homolog
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000
C;Accession: A69284; S45665
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
; Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
; Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: A69284
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-534 <KLE>
A;Cross-references: GB:AE001086; GB:AE000782; NID:g2689409; PIDN:AAB90959.1; PID:g265036
R;Kunow, J.; Linder, D.; Stetter, K.O.; Thauer, R.K.
; Eur. J. Biochem. 223, 503-511, 1994
A;Title: F(420)H(2): Quinone oxidoreductase from Archaeoglobus fulgidus. Characterizatio
A;Reference number: S45665; MUID:94333340; PMID:8055920
A;Accession: S45665
A;Molecule type: protein
A;Residues: 2,'X',4,'X',6-7,'X',9-10,'XX',13-14 <KUN>
A;Note: the authors could not distinguish between glutamate and cysteine for residues 5-
C;Superfamily: glutamate synthase small chain
C;Keywords: oxidoreductase

Query Match      65.4%; Score 34; DB 2; Length 534;
Best Local Similarity 45.5%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
  |||||
Db 119 DKVPHGSHYT 129

RESULT 11
T46758
;Protein: hypothetical 92.4K protein - Streptococcus agalactiae
C;Species: Streptococcus agalactiae
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
R;Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; Luet
; Infect. Immun. 67, 871-878, 1999
A;Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachmen
A;Reference number: Z24091; MUID:99115568; PMID:9916102
A;Accession: T46758
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-822 <SPE>
A;Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624
A;Experimental source: strain R268
C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match      65.4%; Score 34; DB 2; Length 822;
Best Local Similarity 75.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
  |||||
Db 350 VVPHGDHY 357

RESULT 12
F89892
;Protein: carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F89892
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
; ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I

```

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A:Reference number: A99758; PMID:21311952; PMID:11418146
 A:Accession: F89892
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1057 <KUR>
 A:Cross-references: GB:BA000018; PID:gl3701002; PIDN:BA842298.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: pyrAB
 C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 65.4%; Score 34; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 82;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
 ||| |||
 Db 190 EIVSNGLHY 199
 ||| |||

RESULT 13
 E97985
 hypothetical protein spr0909 [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 A:Accession: E97985
 R:Roskins J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; H
 e, K.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 Y. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; PMID:21429245; PMID:11544234
 A:Accession: E97985
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-46 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:gl5458516; GSPDB:GN00174
 C:Genetics:
 A:Gene: spr0909

Query Match 63.5%; Score 33; DB 2; Length 46;
 Best Local Similarity 62.5%; Pred. No. 4.5;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 ||| |||
 Db 21 IVPHGGHY 28
 ||| |||

RESULT 14
 VKLJS1
 trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz
 N:Alternate names: anti-repression trans-activator; art protein; rev protein; trs protein
 C:Species: simian immunodeficiency virus SIVcpz
 A:Note: host Pan troglodytes (Chimpanzee)
 C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C:Accession: S09988
 R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A:Reference number: S09983; PMID:90259077; PMID:2188136
 A:Accession: S09988
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-124 <HUE>
 A:Cross-references: EMBL:X52154; NID:g558866; PIDN:CAA36405.1; PID:g763085
 C:Genetics:
 A:Gene: rev; trs; art
 A:Introns: 27/1

C:Superfamily: AIDS trans-regulatory splicing protein
 C:Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
 ||| |||
 Db 107 ETVFAGGNY 116
 ||| |||

RESULT 15
 D82618
 conserved hypothetical protein XF1950 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82618
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82618
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156 <SIM>
 A:Cross-references: GB:AE004014; GB:AE003849; NID:gg107044; PIDN:AAF84752.1; GSPDB:GN00
 R:Simonsen, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, I
 as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1950

Query Match 63.5%; Score 33; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVPXGXH 9
 ||| |||
 Db 119 BEILPQGVH 127
 ||| |||

Search completed: June 3, 2004, 12:00:02
 Job time : 9 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(Without alignments)
117.693 Million cell updates/sec

File: US-09-909-164-44
Perfect score: 52
Sequence: 1 EEVVPXGXHY 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	69.2	426	1 AROA_VIBPA	Q87qx9 vibrio para
2	36	69.2	1057	1 CARB_STAEP	Q8cpj4 staphylococ
3	35	67.3	1058	1 CARB_FUSNN	Q8cg86 fusobacteri
4	35	67.3	1396	1 VLPF_BPT5	P13390 bacterioph
5	34	65.4	102	1 Y11K_TYDVA	P31619 tobacco yel
6	34	65.4	126	1 TKNK_BOVIN	P08858 bos taurus
7	34	65.4	1057	1 CARB_STAAM	Q99ur5 staphylococ
8	34	65.4	1057	1 CARB_STAAM	P58940 staphylococ
9	33	63.5	124	1 REV_SIVCZ	P17280 chimpanzee
10	33	63.5	460	1 UMEI_YEAST	Q03010 saccharomyc
11	33	63.5	578	1 MDLB_BUCBP	Q89596 buchnera ap
12	33	63.5	743	1 YK47_YEAST	P36148 saccharomyc
13	32	61.5	212	1 MSRA_VIBPA	Q878w6 vibrio para
14	32	61.5	308	1 GAAB_METH	O26806 methanobact
15	32	61.5	363	1 ALFB_SHEEP	P52210 ovis aries
16	32	61.5	573	1 SUOX_DROME	Q9vwp4 drosophila
17	32	61.5	627	1 MUTL_BACSU	P49850 bacillus su
18	32	61.5	735	1 ETRI_BRAOL	Q49230 brassica ol
19	32	61.5	738	1 ETRI_ARATH	P49333 arabidopsis
20	32	61.5	906	1 CENC_MOUSE	P49452 mus musculu
21	32	61.5	1374	1 YQ56_CABEL	Q09541 caenorhabdi
22	32	61.5	1378	1 RON_MOUSE	Q62190 mus musculu
23	32	61.5	1454	1 PTPT_MOUSE	Q99m80 mus musculu
24	32	61.5	1463	1 PTPT_HUMAN	O14522 homo sapien
25	31	59.6	130	1 SZ05_RAT	P97885 rattus norv
26	31	59.6	212	1 MSRA_VIBCH	Q9kp30 vibrio chol
27	31	59.6	264	1 DPHB_METH	O27502 methanobact
28	31	59.6	319	1 YHAI_CRYPA	P10941 cryptoneutr
29	31	59.6	363	1 ALFB_HUMAN	P05062 homo sapien
30	31	59.6	363	1 ALFB_RABIT	P79226 oryctolagus
31	31	59.6	366	1 DDL_CETH	Q8etj6 oceanobacil
32	31	59.6	387	1 THIK_ECO57	Q8x8j4 escherichia
33	31	59.6	387	1 THIK_ECOLI	P21151 escherichia

RESULT 1
AROA_VIBPA STANDARD; PRT; 426 AA.
ID AROA_VIBPA
AC Q87qx9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR VP1020.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yaunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -|- SUBUNIT: Monomer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -|- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005076; BAC59283.1; --
CC HAMAP; MF 00210; --;
DR InterPro; IPR001986; EPSP synth.
DR Pfam; PF00275; EPSP synthase; 1.
DR PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 426 AA; 46094 MW; 373D39CC5BA1F70F CRC64;

Query Match 69.2%; Score 36; DB 1; Length 426;

Best Local Similarity 60.0%; Pred. No. 5.6;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPXGXHY 10

Db 223 EFVIPAGQHY 232

```

RESULT 2
CARB_STAEF
ID CARB_STAEF STANDARD; PRT; 1057 AA.
AC Q8CFJ4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
DE CARB OR SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016746; AAC00476.1; -.
CC HAVAP; MF 01210; -.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPhase_L.
CC InterPro; IPR005479; CPhase_L_D2.
CC InterPro; IPR005480; CPhase_L_D3.
CC InterPro; IPR005481; CPhase_L_N.
CC InterPro; IPR004362; MGS_like.
CC Pfam; PF02787; CPhase_L_chain; 2.
CC Pfam; PF02787; CPhase_L_D2; 2.
CC Pfam; PF02142; CPhase_L_D3; 1.
CC PRINTS; PR00098; CPhase.
CC TIGRPFAMs; TIGR01369; CPhaseII_lrg; 1.
CC PROSITE; PS00866; CPhase_1; 2.
CC PROSITE; PS00867; CPhase_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Complete proteome.
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT NP_BIND 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 284 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 288 298

```

```

FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117391 MW; 89447D7D8DB1CAE59 CRC64;
Query Match 69.2%; Score 36; DB 1; Length 1057;
Best Local Similarity 63.6%; Pred. NO. 14;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGXHYS 11
DB 189 KEVYSNGLHYS 199
RESULT 3
CARB_FUSNN
ID CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
DE CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11899109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Wallunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010554; AAL94625.1; ALT_INIT.
CC HAVAP; MF 01210; -.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPhase_L.
CC InterPro; IPR005479; CPhase_L_D2.
CC InterPro; IPR005480; CPhase_L_D3.
CC InterPro; IPR005481; CPhase_L_N.
CC InterPro; IPR004362; MGS_like.
CC Pfam; PF02786; CPhase_L_chain; 2.
CC Pfam; PF02786; CPhase_L_D2; 2.
CC Pfam; PF02787; CPhase_L_D3; 1.
CC PRINTS; PR00098; CPhase.
CC TIGRPFAMs; TIGR01369; CPhaseII_lrg; 1.

```

R PROSITE; PS00866; CPASE 1; 2.
 W ARGININE BIOSYNTHESIS; Pyrimidine biosynthesis; Ligase; Repeat;
 W ATP-BINDING; Manganese; Complete proteome.
 I DOMAIN 1 401
 I CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 I DOMAIN 402 546
 I OLIGOMERIZATION DOMAIN.
 I DOMAIN 547 929
 I CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 I DOMAIN 930 1058
 I ALLOSTERIC DOMAIN.
 I REPEAT 1 546
 I REPEAT 547 1058
 I NP_BIND 153 210
 I NP_BIND 352 352
 I METAL 284 284
 I METAL 298 298
 I METAL 300 300
 I METAL 820 820
 I METAL 832 832
 I METAL 832 832
 I SEQUENCE 1058 AA; 117451 MW; 5D7037AF77C1E39F CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 23;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Y 2 EVVPXGXHYS 11
 b 190 EIVPGLNYS 199

 RESULT 4
 LTF_BPTS
 ID VITF BPTS STANDARD; PRT; 1396 AA.
 RT 01-JAN-1990 (Rel. 13, Created)
 RT 01-FEB-1996 (Rel. 33, Last sequence update)
 RT 30-MAY-2000 (Rel. 39, Last annotation update)
 RT L-shaped tail fiber protein (LTF protein).
 RT LTF.
 RT Bacteriophage T5.
 RT Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
 RT T5-like viruses.
 RT T5-like viruses.
 RT NCBI_TaxID=10726;
 RT (1) | | | | |
 RT SEQUENCE FROM N.A.
 RT MEDLINE=95309401; PubMed=7789514;
 RT Kallman A.V., Kulshin V.E., Shlyapnikov M.G., Ksenzenko V.N.,
 RT Kryukov V.M.;
 RT "The nucleotide sequence of the bacteriophage T5 ltf gene."
 RT FEBS Lett. 365:46-48(1995).
 RT (2) | | | | |
 RT SEQUENCE FROM N.A.
 RT Kallman A.V.;
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RT (3) | | | | |
 RT PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RT MEDLINE=88289370; PubMed=3267228;
 RT Kallman A.V., Kryukov V.M., Bayev A.A.;
 RT "The nucleotide sequence of bacteriophage T5 DNA at the region
 RT between early and late genes."
 RT Nucleic Acids Res. 16:6230-6230(1988).
 RT -!- FUNCTION: Nonsential protein that mediates binding to the
 RT polyanion O antigen.
 RT -----
 RT This SWISS-PROT entry is copyright. It is produced through a collaboration
 RT between the Swiss Institute of Bioinformatics and the EMBL outstation -
 RT the European Bioinformatics Institute. There are no restrictions on its
 RT use by non-profit institutions as long as its content is in no way
 RT modified and this statement is not removed. Usage by and for commercial
 RT entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 RT or send an email to license@isb-sib.ch).
 RT -----
 DR EMBL; X69460; CAA49220.1; --
 DR EMBL; AJ001191; CAA04591.1; --
 DR PIR; S01982; S01982.

DR PIR; S65934; S36851.
 KW Late protein.
 FT CONFLICT 986 V -> A (IN REF. 2).
 SQ SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 1396;
 Best Local Similarity 50.0%; Pred. No. 31;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPXGXHYS 11
 Db 1360 KTIPTAGDHYS 1369

 RESULT 5
 Y1LK_TYDVA
 ID Y1LK_TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Vi.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Nastrevirus.
 OC NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=92188538; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RA "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants."
 RT Virology 187:633-642(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M81103; AAA47947.1; --
 DR PIR; A42452; A42452.
 DR InterPro; IPR002621; Gemini_mov.
 DR Pfam; PF01708; Gemini_mov; I.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 3.4;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPXGXHYS 11
 Db 7 QVVPGLNYS 16

 RESULT 6
 TKNK_BOVIN
 ID TKNK_BOVIN STANDARD; PRT; 126 AA.
 AC P08858;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurokinin B precursor (NKB) (Neuromedin K).
 GN TAC3 OR NKB.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86313713; PubMed=3462746;
RA Kotani H., Hoshimaru M., Nawa H., Nakanishi S.;
RT "Structure and gene organization of bovine neuromedin K precursor."
RL Proc. Natl. Acad. Sci. U.S.A. 83:7074-7078 (1986).
CC -!- FUNCTION: Tachykinins are active peptides which excite neurons,
CC evoke behavioral responses, are potent vasodilators and
CC secretagogues, and contract (directly or indirectly) many smooth
CC muscles.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the tachykinin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M14351; AAA30723.1;
CC EMBL; M14347; AAA30723.1; JOINED.
CC EMBL; M14348; AAA30723.1; JOINED.
CC EMBL; M14349; AAA30723.1; JOINED.
CC EMBL; M14350; AAA30723.1; JOINED.
CC PIR; A25905; A25905.
CC InterPro; IPR003635; Neurokinin.
CC InterPro; IPR002040; Tachy Neurokinin.
CC Pfam; PF03823; Neurokinin_B; 1.
CC ProDom; PD020370; Neurokinin; 1.
CC ProSite; PS00267; TACHYKININ; 1.
CC Tachykinin; Neurokinin; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 83
FT PEPTIDE 86 95 NEUROKININ B.
FT PROPEP 99 126
FT MOD RES 95 95
SQ SEQUENCE 126 AA; 13871 MW; 446EF433498EC059 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EYVVPXGXH 9
Db 28 EQVVPGGGH 36

RESULT 7
CARB STAAW STANDARD; PRT; 1057 AA.
ID -CARB STAAW STANDARD; PRT; 1057 AA.
AC Q99UR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
DE CARB OR PYRAB OR SAV1203 OR SA1046.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MU50 / ATCC 700699, and N315;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,

```

```

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240 (2001).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AP003361; BAB57365.1;
CC EMBL; AP003132; BAB42298.1;
CC PIR; F89892; F89892.
CC HSP; P00968; ICS0.
CC HAMAP; MF_01210; 1.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; Cbase_L_D2.
CC InterPro; IPR005479; Cbase_L_D2.
CC InterPro; IPR005480; Cbase_L_D3.
CC InterPro; IPR005481; Cbase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF00289; CPSase_L_Chain; 2.
CC Pfam; PF02786; CPSase_L_D2; 2.
CC Pfam; PF02787; CPSase_L_D3; 1.
CC Pfam; PF01442; MGS; 1.
CC PRINTS; PR00098; CPSASE.
CC TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
CC PROSITE; PS00866; CPSASE_1; 2.
CC PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
KW CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 1 401
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1057
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352
FT NP_BIND 284 284
FT METAL 298 298
FT METAL 300 300
FT METAL 820 820
FT METAL 832 832
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EYVVPXGXHS 11
Db 190 EIVSNGLHVS 199

RESULT 8
CARB STAAW STANDARD; PRT; 1057 AA.
ID -CARB STAAW STANDARD; PRT; 1057 AA.
AC P58910;

```

28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
CARB OR PYRAB OR MW1086.
Staphylococcus aureus (strain MW2).
NCBI_TaxID=196620;
[1]
SEQUENCE FROM N.A.
MEDLINE=22040717; PubMed=12044378;
Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.;
"Genome and virulence determinants of high virulence community-acquired MRSA";
Lancet 359:1819-1827(2002).
-1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
-1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
-1- PATHWAY: Arginine biosynthesis.
-1- PATHWAY: Pyrimidine biosynthesis; first step.
-1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
-1- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF004825; BAB94951.1; -;
HAMAP; MF 01210; -; 1.
InterPro; IPR006275; CarA_L_glu.
InterPro; IPR005483; CPase_L_D1.
InterPro; IPR005479; CPase_L_D2.
InterPro; IPR005480; CPase_L_D3.
InterPro; IPR005481; CPase_L_N.
InterPro; IPR004362; MGS like.
Pfam; PF00289; CPase_L_Chain; 2.
Pfam; PF02786; CPase_L_D2; 2.
Pfam; PF02787; CPase_L_D3; 1.
Pfam; PF02142; MGS; 1.
PRINTS; PR00098; CPASE.
TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
PROSITE; PS00866; CPASE_1; 2.
PROSITE; PS00867; CPASE_2; 2.
Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
DOMAIN 930 1057 ALLOSTERIC DOMAIN.
REPEAT 1 546
REPEAT 547 1057
NP_BIND 153 210 ATP (POTENTIAL).
NP_BIND 302 352 ATP (POTENTIAL).
METAL 284 298 MANGANESE 1 (BY SIMILARITY).
METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
METAL 300 300 MANGANESE 2 (BY SIMILARITY).
METAL 820 820 MANGANESE 3 (BY SIMILARITY).
METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SEQUENCE 1057 AA; 117185 MW; D83B09F9BC6F152 CRC64;
Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYHS 11
|:|:|:|:|
DB 190 EIVSNGHYHS 199

RESULT 9
REV_SIVCZ STANDARD; PRT; 124 AA.
AC P17280;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE REV protein (Anti-repression transactivator protein) (ART/TRS).
GN REV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
[1]
SEQUENCE FROM N.A.
MEDLINE=90259077; PubMed=2188136;
Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
"Genetic organization of a chimpanzee lentivirus related to HIV-1.";
Nature 345:356-359(1990).
-1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
CC -1- PTM: Phosphoprotein whose state of phosphorylation is mediated by
CC a specific serine kinase activity present in the nucleus.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X52154; CAA36405.1; -;
PIR; S09988; VKLJGI.
HIV; X52154; REVSCPZ.
InterPro; IPR000635; REV_protein.
Pfam; PF00424; REV; 1.
KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDF65A7B2 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 124;
Best Local Similarity 60.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHYHS 11
|:|:|:|:|
DB 107 ETVPAGNYHS 116

RESULT 10
UWEI_YEAST STANDARD; PRT; 460 AA.
AC Q03010; P87330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Melosis negative regulator UWEI.
GN UWEI OR WTM3 OR YPL139C OR LPI7C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RC STEAIN-A364A;
RA Mallory M.J.; Strich R.;
Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.


```

RN RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth U., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heblung U., Heumann K., Hilbert H., Hallier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Schaefer M.,
RA Scherrens B., Schramm S., Schroeder M., Sedcu A.M., Tettelin H.,
RA Ustretazu L.A., Ushinsky S., Viereckels P., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
CC Nature 387:103-105(1997)
CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10280; AAB40937.1; -
DR EMBL; U43703; AAB68221.1; -
DR PIR; S69046; S69046.
DR GERMOnline; 144121; -
DR TRANSFAC; T04309; -
DR SGD; S0006060; UME1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0037174; F:transcription co-repressor activity; IDA.
DR GO; GO:0040020; P:regulation of meiosis; IGI.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE NEG.
DR PROSITE; PS50082; WD_REPEATS_2; FALSE NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AAGF60448B7BCBA9 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
DB 85 IVPLGLHY 92

RESULT 11
MDLB BUCBP STANDARD; PRT; 578 AA.
AC Q89A96;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

RN RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
CC Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014017; AAO271134.1; -
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR PROSITE; PS50929; ABC_TM1F; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE NEG.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 196 216 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT DOMAIN 339 573 ABC_TRANSPORTER.
FT NP_BIND 373 380 ATP (POTENTIAL).
SQ SEQUENCE 578 AA; 66827 MW; 3B84848CE196ADF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EYVPGXHY 11
DB 550 EYVSGTHY 559

RESULT 12
YK47_YEAST STANDARD; PRT; 743 AA.
AC P36148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCF1-MET1 intergenic region.
DE YK47067W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=97313271; PubMed=9169875;
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth U., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heblung U., Heumann K., Hilbert H., Hallier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Schaefer M.,
RA Scherrens B., Schramm S., Schroeder M., Sedcu A.M., Tettelin H.,
RA Ustretazu L.A., Ushinsky S., Viereckels P., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
CC Nature 387:103-105(1997)
CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10280; AAB40937.1; -
DR EMBL; U43703; AAB68221.1; -
DR PIR; S69046; S69046.
DR GERMOnline; 144121; -
DR TRANSFAC; T04309; -
DR SGD; S0006060; UME1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0037174; F:transcription co-repressor activity; IDA.
DR GO; GO:0040020; P:regulation of meiosis; IGI.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE NEG.
DR PROSITE; PS50082; WD_REPEATS_2; FALSE NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AAGF60448B7BCBA9 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
DB 85 IVPLGLHY 92

RESULT 11
MDLB BUCBP STANDARD; PRT; 578 AA.
AC Q89A96;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Multidrug resistance-like ATP-binding protein mdLB.
GN MDLB OR BAF424.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
CC Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014017; AAO271134.1; -
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR PROSITE; PS50929; ABC_TM1F; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE NEG.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 196 216 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT DOMAIN 339 573 ABC_TRANSPORTER.
FT NP_BIND 373 380 ATP (POTENTIAL).
SQ SEQUENCE 578 AA; 66827 MW; 3B84848CE196ADF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EYVPGXHY 11
DB 550 EYVSGTHY 559

RESULT 12
YK47_YEAST STANDARD; PRT; 743 AA.
AC P36148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCF1-MET1 intergenic region.
DE YK47067W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=97313271; PubMed=9169875;
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Arai H., Aparicio A., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth U., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heblung U., Heumann K., Hilbert H., Hallier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Schaefer M.,
RA Scherrens B., Schramm S., Schroeder M., Sedcu A.M., Tettelin H.,
RA Ustretazu L.A., Ushinsky S., Viereckels P., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
CC Nature 387:103-105(1997)
CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10280; AAB40937.1; -
DR EMBL; U43703; AAB68221.1; -
DR PIR; S69046; S69046.
DR GERMOnline; 144121; -
DR TRANSFAC; T04309; -
DR SGD; S0006060; UME1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0037174; F:transcription co-repressor activity; IDA.
DR GO; GO:0040020; P:regulation of meiosis; IGI.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE NEG.
DR PROSITE; PS50082; WD_REPEATS_2; FALSE NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AAGF60448B7BCBA9 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
DB 85 IVPLGLHY 92

RESULT 11
MDLB BUCBP STANDARD; PRT; 578 AA.
AC Q89A96;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```


Query Match 61.5%; Score 32; DB 1; Length 308;
Best Local Similarity 63.6%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 EVVVPXGXHS 11

Db 216 EVVESGLHES 226

RESULT 15

ALFB SHEEP
ID ALFB SHEEP STANDARD; PRT; 363 AA.
AC P52210; 1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).
GN ALDOB.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Caprinae; Ovis.
CX NCBI_TaxID=9940;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Mesonephros;
RX MEDLINE=3438863; PubMed=8086469;
RA Gnanquinto L., Fahloux E.A., Bezaud J., Servel N., Kirszenbaum M.,
RA Cotinot C.;
RT "Cloning and characterization of a full-length cDNA coding for ovine
RT aldolase B from fetal mesonephros.";
RL Biochim. Biophys. Acta 1219:223-227(1994).
CC -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC phosphate + D-glyceraldehyde 3-phosphate.
CC -!- PATHWAY: Glycolysis; sixth step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- MISCELLANEOUS: In vertebrates, three forms of this ubiquitous
CC glycolytic enzyme are found, aldolase A in muscle, aldolase B in
CC liver and aldolase C in brain.
CC -!- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase
CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z29372; CAA82563.1; -.
DR PIR; S47540; S47540.
DR HSSP; P00883; IADO.
DR InterPro; IPR000741; Aldolase I.
DR Pfam; PF00274; glycolytic enzy; 1.
DR PRODom; PD001128; Aldolase I; 1.
DR PROSITE; PS00158; ALDOLASE_CLASS I; 1.
KW Lyase; Schiff base; Glycolysis; Multigene family.
FT INIT MET 0 BY SIMILARITY.
FT BINDING 55 55 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT BINDING 146 146 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT BINDING 229 229 SCHIFF-BASE WITH DIHYDROXYACETONE-P.
FT ACT_SITE 363 363 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
SQ SEQUENCE 363 AA; 39500 MW; FC8B4566821E2BD CRC64;

Query Match 61.5%; Score 32; DB 1; Length 363;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVVPXGXH 9

Db 189 EVIPDGS 196
Search completed: June 3, 2004, 11:49:55
Job time: 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

WM protein - protein search, using sw model

run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds

(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-44

Perfect score: 52

Sequence: 1 BEVVPXGXHYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	78.8	413	11 Q8K289	Q8K289 mus musculus
2	36	69.2	208	2 Q46486	Q46486 corynebacte
3	36	69.2	1057	16 Q8CFJ4	Q8CFJ4 staphylococ
4	35	67.3	233	10 Q40479	Q40479 nicotiana t
5	35	67.3	237	10 Q9LW50	Q9LW50 nicotiana s
6	35	67.3	285	16 Q9HU6	Q9HU6 rhizobium l
7	35	67.3	308	16 Q9X2E2	Q9X2E2 thermotoga
8	35	67.3	317	9 Q38317	Q38317 lactobacill
9	35	67.3	360	17 Q27679	Q27679 methanobact
10	35	67.3	545	10 Q7XTG3	Q7XTG3 oryza sativ
11	35	67.3	678	12 Q9E1X6	Q9E1X6 cercopithe
12	35	67.3	855	5 Q8IMES	Q8IMES drosophila
13	35	67.3	1028	16 Q8YJ11	Q8YJ11 bruceella me
14	35	67.3	1044	16 Q8DIH0	Q8DIH0 synecococc
15	34	65.4	262	16 Q7WN57	Q7WN57 bordetella
16	34	65.4	262	16 Q7W0Z3	Q7W0Z3 bordetella

Q9rxn9 deinoxococcus
Q9hlh8 thermoplasm
O16912 caenorhabdi
O23966 archaeoglob
Q877y2 streptococc
Q9zbh7 streptococc
Q8e4u1 streptococc
Q8dz81 streptococc
Q8nz82 streptococc
Q8k5g1 streptococc
Q93xv4 streptococc
Q8ib11 plasmodium
Q96yh5 sulfolobus
Q8cyu7 streptococc
Q9xst4 canis famil
Q57489 bacteroides
Q9pc35 xyella fas
Q87d36 xyella fas
Q8i033 bos taurus
Q8i031 bos taurus
O00404 homo sapien
Q87xd7 pseudomonas
Q8ywp1 anabaena sp
Q867a5 tragulus ja
Q8hxy9 bos taurus
Q7vp43 haemophilus
Q9uee9 homo sapien
Q8ign5 drosophila
Q08523 ascaris suu

ALIGNMENTS

RESULT 1

Q8K289 PRELIMINARY; PRT; 413 AA.

AC Q8K289;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to expressed sequence A1987856.
GN TADA3L OR 1110004B19RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC032195; AAH32195.1; -

DR MGD; MGI:1915724; Tada3l

DR GO; GO:0030374; F:ligand-dependent nuclear receptor transcrip. . .; IDA.

DR GO; GO:0005151; F:protein binding; IPI.

SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 78.8%; Score 41; DB 11; Length 413;

Best Local Similarity 63.6%; Pred. No. 2.7;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHYS 11

|||:|:|:|:|

Db 168 BEIPPLGKHYS 178

RESULT 2

Q46486 PRELIMINARY; PRT; 208 AA.

ID Q46486

AC Q46486;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein (Gera).
GN GCRA.
OS Corynebacterium xerosis, and
OS Corynebacterium striatum.
OC Bacteria; Actinobacteria.
OC Corynebacteriaceae; Corynebacteriales.
OC Corynebacteriinae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1725, 43770;
RN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=C.xerosis; STRAIN=M82B;
RC MEDLINE=96117603; PubMed=8559800;
RX Tauch A., Kassing F., Kalinowski J., Puhler A.;
RA "The Corynebacterium xerosis composite transposon Tn5432 consists of
RT two identical insertion sequences, designated IS1249, flanking the
RT erythromycin resistance gene ermC.";
RL Plasmid 34:119-131(1995).
RN [2]
RN SEQUENCE FROM N.A.
RP SPECIES=C.striatum; STRAIN=M82B;
RC MEDLINE=20194806; PubMed=10732668;
RX Tauch A., Kriest S., Kalinowski J., Puhler A.;
RA "The 51,409-bp R-plasmid pT10 from the multi-resistant clinical
RT isolate Corynebacterium striatum M82B is composed of DNA segments
RT initially identified in soil bacteria and in plant, animal, and human
RT pathogens.";
RL Mol. Gen. Genet. 263:1-11(2000).
DR EMBL; U21300; AAC95478.1; -.
DR EMBL; AF024566; AB033590.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 208 AA; 23012 MW; F1504BE1CDE85A6 CRC64;
Query Match 69.2%; Score 36; DB 2; Length 208;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 2 EVVVPXGXHY 11
Db 130 DVPEGKHYA 139
RESULT 3
Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
ID Q8CPJ4
AC Q8CPJ4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Carbamoyl-phosphate synthase large chain.
GN SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 12228;
RC Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016746; AA00476.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006326; F:arginine biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.

DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00289; CPhase_L_chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPhase.
DR TIGRfams; TIGR01369; CPhase_L1g; 1.
DR PROSITE; PS00866; CPhase_L1; 2.
DR PROSITE; PS00867; CPhase_L2; 2.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
KW Complete proteome.
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8B1CAE59 CRC64;
Query Match 69.2%; Score 36; DB 16; Length 1057;
Best Local Similarity 63.6%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EVVVPXGXHY 11
Db 189 KEVWNGLHY 199
RESULT 4
Q40479 PRELIMINARY; PRT; 233 AA.
ID Q40479
AC Q40479
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PRESP-2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BY4; TISSUE=Leaf;
RC MEDLINE=95276459; PubMed=7756828;
RX Ohme-Takagi M., Shinshi H.;
RA "Ethylene-inducible DNA binding proteins that interact with an
RT ethylene responsive element.";
RL Plant Cell 7:173-182(1995).
DR EMBL; D38126; BA007324.1; -.
DR F02590; T02590.
DR HSP; H080337; 2GCC.
DR TRANSFAC; T02654; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PR00367; ETHRSPLEMT.
DR PRODOM; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;
Query Match 67.3%; Score 35; DB 10; Length 233;
Best Local Similarity 60.0%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EVVVPXGXHY 10
Db 90 QAVVPKGRHY 99
RESULT 5
Q9LW50 PRELIMINARY; PRT; 237 AA.
ID Q9LW50
AC Q9LW50
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 Ethylene-responsive element binding factor.
 NSERF2.
 Nicotiana sylvestris (Wood tobacco).
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 lamiales; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4096;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=20399450; PubMed=10945353;
 Kitajima S., Koyama T., Ome-Takagi M., Shinshi H., Sato F.;
 "Characterization of gene expression of NSERFs, transcription factors
 of basic PR genes from Nicotiana sylvestris.";
 Plant Cell Physiol. 41:817-824 (2000).
 EMBL; AB016264; BAA97122.1; -.
 HSP; O80337; 20C.
 GO; GO:0005634; C:nucleus; IEA.
 GO; GO:0003700; P:transcription factor activity; IEA.
 GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 InterPro; IPR001471; IPR.EP.
 Pfam; PF00847; AP2-domain; 1.
 PRINTS; PR00367; ETRSPSELMNT.
 PRODOM; PD011423; TF.EP; 1.
 SMART; SM00380; AP2; 1.
 SEQUENCE 237 AA; 26243 MW; 01BC3EEB51E46298 CRC64;
 Query Match 67.3%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 29;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 1 BEVVPXGXHY 10
 94 QAVVPKGRHY 103
 RESULT 6
 298HU6
 ID Q98HU6 PRELIMINARY; PRT; 285 AA.
 AC Q98HU6
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Cytochrome c1.
 EN MLI2705.
 Rhizobium loti (Mesorhizobium loti).
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=MAFF303099;
 MEDLINE=21082930; PubMed=11214969;
 Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 DNA Res. 7:331-338 (2000).
 EMBL; AF030000; BAB49770.1; -.
 GO; GO:0005746; C:mitochondrial electron transport chain; IEA.
 GO; GO:0005489; P:electron transporter activity; IEA.
 GO; GO:0006118; P:electron transport; IEA.
 InterPro; IPR00345; CytC_heme_BS.
 InterPro; IPR002326; Cyt C1.
 Pfam; PF02167; Cytochrome C1; 1.
 PRINTS; PR00603; CYTOCHROME_C1.
 PROSITE; PS00190; CYTOCHROME_C; 1.
 Complete proteome.
 SEQUENCE 285 AA; 30961 MW; 31D9CDE2711747EE CRC64;

Query Match 67.3%; Score 35; DB 16; Length 285;
 Best Local Similarity 55.6%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 3 VVPXGXHY 11
 194 VIEGTHYN 202
 RESULT 7
 09X2E2
 ID Q9X2E2 PRELIMINARY; PRT; 308 AA.
 AC Q9X2E2
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE FTSH protease activity modulator HFLK.
 EN TW1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NCBI_TaxID=2336;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=MSB8 / DSM 3109;
 MEDLINE=99287316; PubMed=10360571;
 Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima";
 Nature 399:323-329 (1999).
 EMBL; AE001819; AAD36885.1; -.
 PIR; A72207; A72207.
 TIGR; TM1822; -.
 GO; GO:0016020; C:membrane; IEA.
 GO; GO:0008233; F:peptidase activity; IEA.
 InterPro; IPR01107; Band 7.
 InterPro; IPR001972; Stomatin.
 Pfam; PF01145; Band 7; 1.
 PRINTS; PR00721; STOMATIN.
 SMART; SM00244; PHB; 1.
 Protease; Complete proteome.
 SEQUENCE 308 AA; 34778 MW; ADR03603E5101A9D CRC64;
 Query Match 67.3%; Score 35; DB 16; Length 308;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 3 VVPXGXHY 10
 41 VVPXGIRY 48
 RESULT 8
 09X317
 ID Q9X317 PRELIMINARY; PRT; 317 AA.
 AC Q9X317
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Lysin.
 EN LYS.
 OS Lactobacillus bacteriophage phi adh.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 NCBI_TaxID=12417;
 [1]
 SEQUENCE FROM N.A.
 Altermann E.;
 Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=9384014; PubMed=10452953;
RA Altermann E., Klein J., Henrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RL gasseri temperate bacteriophage phi-adh.";
RL Gene 236:333-346(1999).
[3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=95138034; PubMed=7836307;
RA Henrich B., Binischofer B., Blaes U.;
RT "Primary structure and functional analysis of the lysis genes of
RL Lactobacillus gasseri bacteriophage phi-adh.";
RL J. Bacteriol. 177:723-732(1995).
[4]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=93231538; PubMed=8472961;
RA Frenaux C., De Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RL functions from Lactobacillus gasseri temperate bacteriophage phi-
RT adh.";
RL Gene 126:61-66(1993).
[5]
RN RP SEQUENCE FROM N.A.
RX Engel G., Altermann E., Klein J., Henrich B.;
RT "Structure of a genome region of the Lactobacillus gasseri temperate
RL phage phi adh covering a repressor gene and cognate promoters.";
RL Gene 210:67-70(1998).
DR EMBL; AJ131519; CAB52540.1; -.
GO; GO:0003796; F:lysozyme activity; IEA.
GO; GO:0016998; P:cell wall catabolism; IEA.
GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002053; Glyco_hydro_25.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF01183; Glyco_hydro_25; 1.
DR ProDom; PD004620; Glyco_hydro_25; 1.
DR SMART; SM00641; Glyco_25; 1.
DR SMART; SM00287; SH3b; 1.
SQ SEQUENCE 317 AA; 34703 MW; 9FF2715EE43561C7 CRC64;

Query Match 57.3%; Score 35; DB 9; Length 317;
Best Local Similarity 56.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHYA 11
DB 60 VVPNGYHYA 68

RESULT 9
ID O27679 PRELIMINARY; PRT; 360 AA.
AC O27679;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell division protein.
GN MTH1642.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Delta H;
RL MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Alardete T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lamm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Fatwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

```

```

RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000923; AAB86115.1; -.
DR PIR; E69086; E69086.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0000910; P:cytokinesis; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR005140; eRF1_1.
DR InterPro; IPR005141; eRF1_2.
DR InterPro; IPR005142; eRF1_3.
DR InterPro; IPR004405; Pella_3.
DR Pfam; PF03463; eRF1_1; 1.
DR Pfam; PF03464; eRF1_2; 1.
DR Pfam; PF03465; eRF1_3; 1.
DR TIGRfam; TIGR00111; pelota; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 360 AA; 40814 MW; 2A000CB4B3CEFA69 CRC64;

Query Match 67.3%; Score 35; DB 17; Length 360;
Best Local Similarity 45.5%; Pred. No. 46;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPGXGHYS 11
DB 98 EDLVPNGSHHT 108

RESULT 10
ID Q7XTG3 PRELIMINARY; PRT; 545 AA.
AC Q7XTG3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Q99121_12.4 protein.
GN Q99121_12.4.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.-Y., Ren S.-X., Lv G., Lin W., Gu W.-Q., Zhu G.-F., Tu Y.-F.,
RA Jia J., Yin H.-F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.-Y., Ding C.-W.,
RA Shao Y., Sun Y., Hu Q.-P., Zhang X.-L., Zhang W., Wang L.-J., Wang C.-W.,
RA Sheng H.-H., Gu J.-L., Chen S.-T., Ni L., Zhu F.-H., Han B., Feng Q.,
RA Huang Y.-C., Li Y., Zhu J.-J., Zhao Q., Hu X., Liu Y.-L., Mu J., Yu Z.,
RA Chen L., Fan D.-L., Wang Q.-J., Zhang L., Lu Y.-Q., Yu S.-L., Liu X.-H.,
RA Lu T.-T., Zhang Y.-J., Lu Y., Li C.-L., Zhang Y., Hu H., Jia P.-X.,
RA Qian Y.-M., Ying K., Zhou B., Chen Z.-H., Hao P., Zhang L., Wu M.,
RA Zhang R.-Q., Guan J.-P., Hong G.-F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL060453; CAB01515.1; -.
SQ SEQUENCE 545 AA; 59078 MW; 4629A5D1DA538692 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 545;
Best Local Similarity 55.6%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHYA 11
DB 433 IVPSGDHYA 441

RESULT 11
ID Q9E1X6 PRELIMINARY; PRT; 678 AA.
AC Q9E1X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE Hypothetical protein.
 DS Cercopithecine herpesvirus 7.
 DC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 DX Alphaherpesvirinae; Varicellovirus.
 [1]
 NCBI_TaxID=35245;
 RP SEQUENCE FROM N.A.
 RA Gray W.L., Scarnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
 "Complete sequence of the Simian Varicella Virus Genome.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275348; FAG27217.1;
 DR GO; GO:0019012; C:virion; IEA.
 DR GO; GO:0006323; P:DNA packaging; IEA.
 DR InterPro; IPR007640; Herpes_UL17.
 DR Pfam; PF04559; Herpes_UL17; 1.
 KW Hypothetical protein_UL17; 1.
 SQ SEQUENCE 678 AA; 75850 MW; AL7B09E30512FE3C CRC64;
 Query Match 67.3%; Score 35; DB 12; Length 678;
 Best Local Similarity 50.0%; Pred. No. 90;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 DY 1 BEVVPXGXHY 10
 ||:|:|
 DB 147 EIIIPKGRY 156
 RESULT 12
 Q8IME6 PRELIMINARY; PRT; 855 AA.
 AC Q8IME6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG30437-PC.
 GN CG30437 OR CG10398 OR CG10408.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 DX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherz S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
 RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glocker A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.H., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Smith H.O.,
 RA Zheng X.H., Zhong F.N., Zhong B.W., Rubin G.M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Dou L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Fargas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Furi V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003786; AAN16124.1; -
 DR FlyBase; FBgn0050437; CG30437.
 DR GO; GO:0005007; F:copper ion binding; IEA.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR008972; Cupredoxin.
 DR InterPro; IPR002355; Cu_ox_copper_BS.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 SQ SEQUENCE 855 AA; 94532 MW; 39BD5A516D6312DB CRC64;
 Query Match 67.3%; Score 35; DB 5; Length 855;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 BEVVPXGXH 9
 ||:|:|
 DB 657 DEVVPSGDH 665
 RESULT 13
 Q8VJ11 PRELIMINARY; PRT; 1028 AA.
 ID Q8VJ11
 AC Q8VJ11
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ATP-dependent DNA helicase.


```

GN BMEI0275.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
CX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haelkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009470; AAL51457.1; -.
DR PIR; AF3286; AF3286.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SW00450; HELIC_C; 1.
DR PROSITE; PS00587; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1028;
Best Local Similarity 54.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHS 11
Db 76 EKIVPPGARYS 86
: : : : :
: : : : :

RESULT 14
QSDIHO PRELIMINARY; PRT; 1044 AA.
AC QSDIHO;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1616.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
CX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AF005374; BAC09170.1; -.
DR GO; GO:0015021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR001036; AcrIIV_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; AcrIIV; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.

```

```

DR TIGRFAMS; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHS 11
Db 843 EEVLPNGIGYS 853
: : : : :
: : : : :

RESULT 15
Q7WNB7 PRELIMINARY; PRT; 262 AA.
AC Q7WNB7;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN B81123.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
CX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebatia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,
RA Cerdano-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltham T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsis K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640440; CAE31621.1; -.
KW Complete proteome.
SQ SEQUENCE 262 AA; 28907 MW; B3CA29331CB776B2 CRC64;

Query Match 65.4%; Score 34; DB 16; Length 262;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXH 9
Db 182 QEVVPYQGH 190
: : : : :
: : : : :

Search completed: June 3, 2004, 11:57:34
Job time : 29.8667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

WM protein - protein search, using sw model

run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 EVVVPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	54.0	11	5 ABB80548	Hepatitis
2	47	94.0	11	5 ABB80547	Hepatitis
3	47	94.0	11	5 ABB80556	Hepatitis
4	47	94.0	11	5 ABB80557	Hepatitis
5	47	94.0	11	5 ABB80551	Hepatitis
6	46	92.0	11	5 ABB80534	Hepatitis
7	46	92.0	11	5 ABB80561	Hepatitis
8	46	92.0	11	5 ABB80542	Hepatitis
9	46	92.0	11	5 ABB80543	Hepatitis
10	46	92.0	11	5 ABB80524	Hepatitis
11	46	92.0	11	5 ABB80533	Hepatitis
12	46	92.0	11	5 ABB80529	Hepatitis
13	46	92.0	11	5 ABB80528	Hepatitis
14	46	92.0	11	5 ABB80538	Hepatitis
15	46	92.0	11	5 ABB80562	Hepatitis
16	41	82.0	11	5 ABB80549	Hepatitis
17	41	82.0	11	5 ABB80544	Hepatitis
18	41	82.0	11	5 ABB80553	Hepatitis
19	41	82.0	11	5 ABB80552	Hepatitis
20	41	82.0	11	5 ABB80545	Hepatitis
21	40	80.0	11	5 ABB80525	Hepatitis
22	40	80.0	11	5 ABB80521	Hepatitis
23	40	80.0	11	5 ABB80522	Hepatitis
24	40	80.0	11	5 ABB80536	Hepatitis
25	40	80.0	11	5 ABB80566	Hepatitis

26	40	80.0	11	5 ABB80546	Hepatitis
27	40	80.0	11	5 ABB80563	Hepatitis
28	40	80.0	11	5 ABB80565	Hepatitis
29	40	80.0	11	5 ABB80535	Hepatitis
30	40	80.0	11	5 ABB80567	Hepatitis
31	40	80.0	11	5 ABB80540	Hepatitis
32	40	80.0	11	5 ABB80554	Hepatitis
33	40	80.0	11	5 ABB80559	Hepatitis
34	40	80.0	11	5 ABB80526	Hepatitis
35	40	80.0	11	5 ABB80530	Hepatitis
36	40	80.0	11	5 ABB80539	Hepatitis
37	40	80.0	11	5 ABB80550	Hepatitis
38	40	80.0	11	5 ABB80555	Hepatitis
39	40	80.0	11	5 ABB80564	Hepatitis
40	40	80.0	11	5 ABB80568	Hepatitis
41	39	78.0	11	5 ABB80523	Hepatitis
42	39	78.0	11	5 ABB80558	Hepatitis
43	39	78.0	11	5 ABB80537	Hepatitis
44	39	78.0	11	5 ABB80560	Hepatitis
45	39	78.0	11	5 ABB80527	Hepatitis

ALIGNMENTS

RESULT 1					
ABB80548					
ID	ABB80548 standard; peptide; 11 AA.				
XX					
AC	ABB80548;				
XX					
DT	08-OCT-2002 (first entry)				
XX					
DR	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.				
XX					
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.				
XX					
OS	Synthetic.				
XX					
PH	Key Location/Qualifiers				
FT	Modified-site 1 /note= "N-terminal acetyl"				
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"				
FT	Misc-difference 9 /note= "D-form residue"				
FT	Modified-site 11 /note= "C-terminal amide"				
XX	WO200208251-A2.				
XX	31-JAN-2002.				
XX	19-JUL-2001; 2001WO-US023169.				
XX	21-JUL-2000; 2000US-0220101P.				
XX	(CORV-) CORVAS INT INC.				
XX	Lim-Wilby M, Levy OE, Brunck TK;				
XX	WPI; 2002-361643/39.				
XX	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.				
XX	Claim 17; Page 65; 69pp; English.				
XX	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the				

CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
| | | | | | | | | |
Db 1 EEVVPXGTDYS 11

RESULT 2
ABB80547
ID ABB80547 standard; peptide; 11 AA.

XX AC ABB80547;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.
XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus

XX Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.012;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXDYS 11
| | | | | | | | | |
Db 1 EEVVPXGTDYS 11

RESULT 3
ABB80556

ID ABB80556 standard; peptide; 11 AA.

XX AC ABB80556;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #36.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.

XX Synthetic.

XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus

XX Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
| | | | | | | | | |
Db 1 EEVVPXGTDYS 11

RESULT 4

XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.
DE	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX	virucide.
KW	Synthetic.
OS	Key
XX	Location/Qualifiers
FH	Modified-site 1
FT	/note= "N-terminal acetyl"
FT	6
FT	Modified-site
FT	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 9
FT	/note= "D-form residue"
FT	11
FT	Modified-site
FT	/note= "C-terminal amide"
XX	WO200208251-A2.
PX	31-JAN-2002.
XX	19-JUL-2001; 2001WO-US023169.
PX	21-JUL-2000; 2000US-022010IP.
PR	(CORV-) CORVAS INT INC.
XX	Lim-Wilby M, Levy OE, Brunck TK;
PA	WPI; 2002-361643/39.
XX	Noel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
PT	Claim 17; Page 65; 69pp; English.
PS	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides have invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
CC	Sequence 11 AA;
XX	Query Match 94.0%; Score 47; DB 5; Length 11;
XX	Best Local Similarity 90.9%; Pred. No. 0.012;
XX	Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY	1 BEVPXGXDYS 11
Db	1 BEVPXGSDYS 11
RESULT 6	
ABB80534	ID ABB80534 standard; peptide; 11 AA.
XX	AC ABB80534;
XX	DT 08-OCT-2002 (first entry)
XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #14.
DE	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX	virucide.
KW	Synthetic.
OS	Key
XX	Location/Qualifiers
FH	Modified-site 1
FT	/note= "N-terminal acetyl"
FT	6
FT	Modified-site
FT	/note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 9
FT	/note= "D-form residue"
FT	11
FT	Modified-site
FT	/note= "C-terminal amide"
XX	WO200208251-A2.
PX	31-JAN-2002.
XX	19-JUL-2001; 2001WO-US023169.
PX	21-JUL-2000; 2000US-022010IP.
PR	(CORV-) CORVAS INT INC.
XX	Lim-Wilby M, Levy OE, Brunck TK;
PA	WPI; 2002-361643/39.
XX	Noel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
PT	Claim 17; Page 65; 69pp; English.
PS	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus
CC	Sequence 11 AA;
XX	Query Match 94.0%; Score 47; DB 5; Length 11;
XX	Best Local Similarity 90.9%; Pred. No. 0.012;
XX	Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 BEVPXGXDYS 11
b	1 BEVPXGSDYS 11
RESULT 5	
ABB80551	D ABB80551 standard; peptide; 11 AA.
XX	C ABB80551;
X	08-OCT-2002 (first entry)

XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
XX
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXQDYS 11
DB 1 EEVVPXGXQDYS 11
RESULT 9
ABB80543
ID ABB80543 standard; peptide; 11 AA.
XX
XX ABB80543;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 9 /note= "D-form residue"
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
XX
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXQDYS 11
DB 1 EEVVPXGXQDYS 11
RESULT 10
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
XX ABB80524;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX Misc-difference 9 /note= "D-form residue"
XX Modified-site 11 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.02;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXDYS 11

Db 1 EEVVPXGMDYS 11

RESULT 11

ABB80533

ID ABB80533 standard; peptide; 11 AA.

AC ABB80533;

XX 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #13.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

XX Synthetic.

OS Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Modified-site 11 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus

PT protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.02;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGXDYS 11

Db 1 EEVVPXGMDYS 11

RESULT 12

ABB80529

ID ABB80529 standard; peptide; 11 AA.

XX ABB80529;

XX 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

XX Synthetic.

OS Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Modified-site 7" residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C virus

PT protease.

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.02;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 BEVVPXGXDYS 11
 1 BEVVPXGWDYS 11

RESULT 13

ABB80528
 ID ABB80528 standard; peptide; 11 AA.

AC ABB80528;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

GW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

GW virucide.

DS Synthetic.

FX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT residue 7"

FT Misc-difference 8

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

FX WO200208251-A2.

FX 31-JAN-2002.

FX 19-JUL-2001; 2001WO-US023169.

FX 21-JUL-2000; 2000US-0220101P.

FX (CORV-) CORVAS INT INC.

FX Lim-Wilby M, Levy OE, Brunck TK;

FX WPI; 2002-361643/39.

FX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

FX Claim 17; Page 64; 69pp; English.

FX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

FX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11

Db 1 BEVVPXGWDYS 11

RESULT 14

ABB80538

ID ABB80538 standard; peptide; 11 AA.

AC ABB80538;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.

GW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

GW virucide.

DS Synthetic.

FX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

FX WO200208251-A2.

FX 31-JAN-2002.

FX 19-JUL-2001; 2001WO-US023169.

FX 21-JUL-2000; 2000US-0220101P.

FX (CORV-) CORVAS INT INC.

FX Lim-Wilby M, Levy OE, Brunck TK;

FX WPI; 2002-361643/39.

FX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

FX Claim 17; Page 64; 69pp; English.

FX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus

FX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11

Db 1 BEVVPXGWDYS 11

RESULT 15

ABB80562

ID ABB80562 standard; peptide; 11 AA.

AC ABB80562;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

XX

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT Modified-site 6 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT Modified-site 8 residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 PN
 XX
 XX
 PD 31-JAN-2002.
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-Wilby M, Levy OE, Brunck TK;
 PI
 PI WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 92.0%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 BEVVPXGXDYS 11
 |||||
 Db 1 BEVVPXGMDYS 11
 Search completed: June 3, 2004, 11:48:24
 Job time : 45.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run On: June 3, 2004, 11:36:47 ; Search time 11.7333 seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 EVVFXGXDXS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	74.0	181	4	US-09-134-000C-4848
2	33	66.0	140	3	US-08-569-147-76
3	33	66.0	140	3	US-08-569-147-82
4	32	64.0	139	4	US-09-134-001C-5124
5	32	64.0	301	4	US-08-830-217-14
6	32	64.0	3472	4	US-09-408-020-4
7	31	62.0	303	4	US-09-134-000C-4318
8	31	62.0	622	2	US-08-459-146-2
9	31	62.0	622	2	US-08-459-065-2
10	31	62.0	739	3	US-09-413-814-86
11	30.5	61.0	669	4	US-09-617-594A-2
12	30	60.0	59	4	US-08-963-851-14
13	30	60.0	120	4	US-09-621-976-4604
14	30	60.0	121	4	US-09-152-060-68
15	30	60.0	121	4	US-09-152-060-85
16	30	60.0	122	2	US-08-879-995A-1
17	30	60.0	122	3	US-09-215-096-1
18	30	60.0	217	4	US-09-543-681A-5172
19	30	60.0	231	3	US-08-926-842B-20
20	30	60.0	240	3	US-08-926-842B-21
21	30	60.0	241	3	US-08-834-776A-2
22	30	60.0	341	3	US-08-853-948B-4
23	30	60.0	348	3	US-08-853-948B-5
24	30	60.0	368	4	US-09-697-367-24
25	30	60.0	421	3	US-09-093-448-4
26	30	60.0	421	3	US-09-813-555-4
27	30	60.0	421	4	US-09-523-263B-18

28 30 60.0 425 4 US-09-252-991A-25088 Sequence 25088, A
29 30 60.0 478 2 US-09-040-799-3 Sequence 3, Appli
30 30 60.0 478 3 US-09-093-448-1 Sequence 1, Appli
31 30 60.0 478 3 US-09-093-448-2 Sequence 2, Appli
32 30 60.0 478 3 US-09-093-448-3 Sequence 3, Appli
33 30 60.0 478 4 US-09-813-555-1 Sequence 1, Appli
34 30 60.0 478 4 US-09-813-555-2 Sequence 2, Appli
35 30 60.0 478 4 US-09-813-555-3 Sequence 3, Appli
36 30 60.0 478 4 US-09-813-555-4 Sequence 4, Appli
37 30 60.0 478 4 US-09-523-263B-4 Sequence 16, Appli
38 30 60.0 478 4 US-09-523-263B-17 Sequence 17, Appli
39 30 60.0 480 4 US-09-252-991A-22164 Sequence 22164, A
40 30 60.0 864 4 US-09-810-268-3 Sequence 3, Appli
41 30 60.0 1049 4 US-09-394-272-10 Sequence 10, Appli
42 30 60.0 1068 2 US-08-429-054A-11 Sequence 11, Appli
43 30 60.0 1068 2 US-08-718-777-7 Sequence 7, Appli
44 30 60.0 1068 2 US-09-051-341-7 Sequence 7, Appli
45 30 60.0 1068 4 US-09-394-272-8 Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-134-000C-4848
; Sequence 4848, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4848
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 74.0%; Score 37; DB 4; Length 181;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVFXGXDX 10
Db 145 EVVFTSEYD 154

RESULT 2
US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESS: No. 6180377ris, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-76

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDX 10
DB 122 VVPTGFY 129

RESULT 3
US-08-569-147-82
Sequence 82, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 6180377ris, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-82

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDX 10
DB 122 VVPTGFY 129

RESULT 4
US-09-134-001C-5124
Sequence 5124, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5124
LENGTH: 139
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5124

Query Match 64.0%; Score 32; DB 4; Length 139;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDXS 11
DB 32 IVFFGHDYN 40

RESULT 5
US-09-830-217-14
Sequence 14, Application US/09830217
Patent No. 6521441
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
FILE REFERENCE: PB461PCT
CURRENT APPLICATION NUMBER: US/09/830,217
CURRENT FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: PCT/US99/06199
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/080,296
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/084,674
PRIOR FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 301
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-830-217-14

Query Match 64.0%; Score 32; DB 4; Length 301;
Best Local Similarity 55.6%; Pred. No. 92;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDXS 11
DB 195 IVFFGHDYN 203

RESULT 6
US-09-408-020-4
Sequence 4, Application US/09408020
Patent No. 6632937

GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCRP.002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3472
TYPE: PRT
ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 64.0%; Score 32; DB 4; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
|:|:|:|:|:
DB 2294 EDVIRGISFS 2304

RESULT 7
US-09-134-000C-4318
Sequence 4318, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4318
LENGTH: 303
TYPE: PRT
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (2)-(2)
OTHER INFORMATION: Amino acid 2 is Xaa wherein Xaa = any amino acid.
US-09-134-000C-4318

Query Match 62.0%; Score 31; DB 4; Length 303;
Best Local Similarity 40.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
|:|:|:|:|:
DB 222 EQITPTGIEY 231

RESULT 8
US-08-459-146-2
Sequence 2, Application US/08459146
Patent No. 5866405
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,146
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R.
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryptonectria)
ORGANISM: parasitica)
STRAIN: BP713
US-08-459-146-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
|:|:|:|:|:
DB 31 BEVVPAG 37

RESULT 9
US-08-459-065-2
Sequence 2, Application US/08459065
Patent No. 5882642
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptosporidia
; ORGANISM: Parasitica)
; STRAIN: EF713
; US-08-459-065-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
|||
DB 31 BEVVPAG 37

RESULT 10

US-09-413-814-86
; Sequence 86, Application US/09413814
; Patent No. 6225054

; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 93/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent ver. 2.1
; SEQ ID NO 86
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-413-814-86

Query Match 62.0%; Score 31; DB 3; Length 739;
Best Local Similarity 62.5%; Pred. No. 4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGXGDXYS 11
|||
DB 663 IPLGGDXYS 670

RESULT 11

US-09-617-594A-2

; Sequence 2, Application US/09617594A
; Patent No. 6541458
; GENERAL INFORMATION:
; APPLICANT: Audonnet, et al.
; TITLE OF INVENTION: FELINE CALICIVIRUS GENES AND VACCINES IN PARTICULAR RECOMBINANT
; FILE REFERENCE: 454313-3151-1
; CURRENT APPLICATION NUMBER: US/09/617,594A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/193,332
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: France 00 01761
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: France 99 09421
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Feline calicivirus
; US-09-617-594A-2

Query Match 61.0%; Score 30.5; DB 4; Length 669;
Best Local Similarity 63.6%; Pred. No. 4.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 BEVVPXGDXYS 11
|||
DB 428 EELTPAG-DYS 437

RESULT 12

US-08-963-851-14
; Sequence 14, Application US/08963851
; Patent No. 6300116

; GENERAL INFORMATION:
; APPLICANT: VAN DER OSTEN, CLAUS
; APPLICANT: HANKER, TORDEN
; APPLICANT: ANDERSEN, CARSTEN
; APPLICANT: BAUDITZ, PETER
; APPLICANT: HANSEN, PETER KAMP
; TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
; FILE REFERENCE: 4946,200-US
; CURRENT APPLICATION NUMBER: US/08/963,851
; CURRENT FILING DATE: 1997-11-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-08-963-851-14

Query Match 60.0%; Score 30; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGDXYS 11
|||
DB 38 EKHPPGSLYS 48

RESULT 13

US-09-621-976-4604
; Sequence 4604, Application US/09621976
; Patent No. 6639063

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 4604
 ; LENGTH: 120
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-621-976-4604

Query Match 60.0%; Score 30; DB 4; Length 120;
 Best Local Similarity 50.0%; Pred. No. 82;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGSDYS 11
 Db 37 EILPSSDRS 46

RESULT 14
 US-09-152-060-68
 ; Sequence 68, Application US/09152060
 ; Patent No. 6448230
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 28 Human Secreted Proteins
 ; FILE REFERENCE: P2003P1.US
 ; CURRENT FILING DATE: 1998-09-11
 ; EARLIER APPLICATION NUMBER: PCT/US98/04858
 ; EARLIER FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: 60/040,762
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/040,710
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/050,934
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,100
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,357
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,189
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/057,765
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/048,970
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/068,368
 ; EARLIER FILING DATE: 1997-12-19
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 68
 ; LENGTH: 121
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-152-060-68

Query Match 60.0%; Score 30; DB 4; Length 121;
 Best Local Similarity 85.7%; Pred. No. 83;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXG 7
 Db 28 EVVVPFG 34

RESULT 15
 US-09-152-060-85
 ; Sequence 85, Application US/09152060
 ; Patent No. 6448230
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 28 Human Secreted Proteins
 ; FILE REFERENCE: P2003P1.US

; CURRENT APPLICATION NUMBER: US/09/152,060
 ; CURRENT FILING DATE: 1998-09-11
 ; EARLIER APPLICATION NUMBER: PCT/US98/04858
 ; EARLIER FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: 60/040,762
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/040,710
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/050,934
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,100
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,357
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,189
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/057,765
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/048,970
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/068,368
 ; EARLIER FILING DATE: 1997-12-19
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 85
 ; LENGTH: 121
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: SITE
 ; LOCATION: (67)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (89)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-152-060-85

Query Match 60.0%; Score 30; DB 4; Length 121;
 Best Local Similarity 85.7%; Pred. No. 83;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
 Db 28 BEVVPFG 34

Search completed: June 3, 2004, 12:03:09
 Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 BEVVPXGXDS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 28138677 residues 1155919
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	94.0	11	12	US-09-909-164-31
2	47	94.0	11	12	US-09-909-164-32
3	47	94.0	11	12	US-09-909-164-35
4	47	94.0	11	12	US-09-909-164-40
5	47	94.0	11	12	US-09-909-164-41
6	46	92.0	11	12	US-09-909-164-8
7	46	92.0	11	12	US-09-909-164-12
8	46	92.0	11	12	US-09-909-164-13
9	46	92.0	11	12	US-09-909-164-17
10	46	92.0	11	12	US-09-909-164-18
11	46	92.0	11	12	US-09-909-164-22
12	46	92.0	11	12	US-09-909-164-26
13	46	92.0	11	12	US-09-909-164-27
14	46	92.0	11	12	US-09-909-164-45
15	46	92.0	11	12	US-09-909-164-46

16	41	82.0	11	12	US-09-909-164-28	Sequence 28, Appl
17	41	82.0	11	12	US-09-909-164-29	Sequence 29, Appl
18	41	82.0	11	12	US-09-909-164-33	Sequence 33, Appl
19	41	82.0	11	12	US-09-909-164-36	Sequence 36, Appl
20	41	82.0	11	12	US-09-909-164-37	Sequence 37, Appl
21	40	80.0	11	12	US-09-909-164-5	Sequence 5, Appl
22	40	80.0	11	12	US-09-909-164-6	Sequence 6, Appl
23	40	80.0	11	12	US-09-909-164-9	Sequence 9, Appl
24	40	80.0	11	12	US-09-909-164-10	Sequence 10, Appl
25	40	80.0	11	12	US-09-909-164-14	Sequence 14, Appl
26	40	80.0	11	12	US-09-909-164-19	Sequence 19, Appl
27	40	80.0	11	12	US-09-909-164-20	Sequence 20, Appl
28	40	80.0	11	12	US-09-909-164-23	Sequence 23, Appl
29	40	80.0	11	12	US-09-909-164-24	Sequence 24, Appl
30	40	80.0	11	12	US-09-909-164-30	Sequence 30, Appl
31	40	80.0	11	12	US-09-909-164-34	Sequence 34, Appl
32	40	80.0	11	12	US-09-909-164-38	Sequence 38, Appl
33	40	80.0	11	12	US-09-909-164-39	Sequence 39, Appl
34	40	80.0	11	12	US-09-909-164-43	Sequence 43, Appl
35	40	80.0	11	12	US-09-909-164-47	Sequence 47, Appl
36	40	80.0	11	12	US-09-909-164-48	Sequence 48, Appl
37	40	80.0	11	12	US-09-909-164-49	Sequence 49, Appl
38	40	80.0	11	12	US-09-909-164-50	Sequence 50, Appl
39	40	80.0	11	12	US-09-909-164-51	Sequence 51, Appl
40	40	80.0	11	12	US-09-909-164-52	Sequence 52, Appl
41	39	78.0	11	12	US-09-909-164-7	Sequence 7, Appl
42	39	78.0	11	12	US-09-909-164-11	Sequence 11, Appl
43	39	78.0	11	12	US-09-909-164-15	Sequence 15, Appl
44	39	78.0	11	12	US-09-909-164-16	Sequence 16, Appl
45	39	78.0	11	12	US-09-909-164-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-909-164-31
Sequence 31, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 31
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)-(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)-(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)-(6)
OTHER INFORMATION: norvaline-(CO)
US-09-909-164-31

Query Match 94.0% Score 47; DB 12; Length 11;

; FILE REFERENCE: IN01192-US


```
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-18

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
Db 1 BEVVPXGQDYS 11

RESULT 11
US-09-909-164-22
; Sequence 22, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-26

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
Db 1 BEVVPXGQDYS 11

RESULT 13
US-09-909-164-27
; Sequence 27, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-22

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
Db 1 BEVVPXGQDYS 11

RESULT 12
US-09-909-164-26
; Sequence 26, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
```

OTHER INFORMATION: AMIDATION

NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(9)
OTHER INFORMATION: D-amino acids
US-09-909-164-27

Query Match 92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11

DB 1 EEVVPXGQDYS 11

RESULT 14

US-09-909-164-45
Sequence 45; Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 45
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: Met (O)
US-09-909-164-45

Query Match 92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11

DB 1 EEVVPXGXDYS 11

RESULT 15

US-09-909-164-46
Sequence 46; Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 46
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline-(CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acids
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: Met (O)
US-09-909-164-46

Query Match 92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11

DB 1 EEVVPXGXDYS 11

Search completed: June 3, 2004, 12:57:17
Job time : 34.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-45

Perfect score: 50

Sequence: 1 EVVFPXGXDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : FIR 78.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	363	2 D69551	conserved hypothet
2	35	70.0	290	2 AG3104	6-O-methylguanine-
3	35	70.0	290	2 D98182	O6-methylguanine-D
4	35	70.0	587	2 P81138	succinate dehydrog
5	35	70.0	906	2 T48898	disease resistance
6	35	70.0	908	2 T48899	disease resistance
7	34	68.0	102	2 A42452	V1 protein - tobac
8	34	68.0	156	2 S54619	hypothetical prote
9	34	68.0	247	2 A96001	conserved hypothet
10	34	68.0	257	2 A96546	unknown protein (i
11	34	68.0	394	2 P82491	ferrisiderophore r
12	34	68.0	433	2 P87660	peptidoglycan-bind
13	34	68.0	2747	2 B49132	fat facets (faf) s
14	33	66.0	124	1 VKLJ51	trans-regulatory s
15	33	66.0	165	2 AG1272	thiol peroxidases
16	33	66.0	165	2 AH1635	thiol peroxidases
17	33	66.0	196	2 AD0454	conserved hypothet
18	33	66.0	225	2 S57810	hypothetical prote
19	33	66.0	327	2 S40753	hypothetical prote
20	33	66.0	421	1 D8RTCM	acyl-CoA dehydrog
21	33	66.0	440	2 H72784	probable alkaline
22	33	66.0	1028	2 AF3286	Arp-dependent DNA
23	33	66.0	1088	2 D82246	probable chitinase
24	33	66.0	1150	2 T20173	hypothetical prote
25	32	64.0	99	2 S00210	plastocyanin b - L
26	32	64.0	155	2 S38255	plastocyanin precu
27	32	64.0	168	2 S58208	plastocyanin b pre
28	32	64.0	196	2 A10931	conserved hypothet
29	32	64.0	301	2 P89957	hypothetical prote

ALIGNMENTS

RESULT 1

D69551 conserved hypothetical protein AF2411 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C/Accession: D69551

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Wese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: D69551

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-363 <KLE>

A/Cross-references: GB:AE001109; GB:AE000782; NID:G2689432; PIDN:AB91255.1; PID:G265068

Query Match 72.0%; Score 36; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVFPXGXDYS 11
| | | | | | | | | |
Db 120 ENIVPVGIDFS 130

RESULT 2

AG3104

6-O-methylguanine-DNA methyltransferase [imported] - Agrobacterium tumefaciens (strain C

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AG3104

R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L

; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AG3104

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-290 <KUR>

A/Cross-references: GB:AE008689; PIDN:AAL45253.1; PID:G17742937; GSPDB:GN00187

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: ada

A/Map position: linear chromosome

```

Query Match      70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
   |||
Db 9 EDITPIGSDY 18

RESULT 3
D98182
O6-methylguanine-DNA methyltransferase PA2118 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: D98182
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, D.; Doughty, D.; Scott, C.; Ippas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A/Accession: D98182
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-290 <KUR>
A/Cross-references: GB:AE007870; PIDN:AAK89882.1; PID:G15158766; GSPDB:GN00170
C/Genetics:
A/Gene: AGR_L818
A/Map position: linear chromosome

Query Match      70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
   |||
Db 9 EDITPIGSDY 18

RESULT 4
F81138
succinate dehydrogenase, flavoprotein chain NMB0950 [imported] - Neisseria meningitidis C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C/Accession: F81138
R/Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: F81138
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-587 <TET>
A/Cross-references: GB:AE002046; GB:AE002098; NID:G7226185; PIDN:AAF41356.1; PID:G722618
A/Experimental source: serogroup B, strain MC58
C/Genetics:
A/Gene: NMB0950
C/Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; F01138

Query Match      70.0%; Score 35; DB 2; Length 587;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 3; Gaps 0;

QY 1 EEVVPXGXDY 10
   |||
Db 366 EVVVPQGEDY 375

RESULT 5
T48898
disease resistance protein RPP8 [validated] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C/Accession: T48898
R/McDowell, J.M.; Dhandaaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dang Plant Cell 10, 1861-1874, 1998
A/Title: Intragenic recombination and diversifying selection contribute to the evolution of the RPP8 gene in Arabidopsis thaliana
A/Reference number: Z24999; MUID:99030193; PMID:9811794
A/Accession: T48898
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-906 <MCD>
A/Cross-references: EMBL:AF089710; NID:G3928861; PIDN:AAK83165.1; PID:G3928862
A/Experimental source: Landsberg erecta
C/Genetics:
A/Gene: RPP8
A/Introns: 293/1; 342/1
C/Function:
A/Description: promotes resistance to Peronospora parasitica

Query Match      70.0%; Score 35; DB 2; Length 906;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
   |||
Db 881 EKLVPQGEDY 890

RESULT 6
T48899
disease resistance protein rpp8 [similarity] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C/Accession: T48899
R/McDowell, J.M.; Dhandaaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dang Plant Cell 10, 1861-1874, 1998
A/Title: Intragenic recombination and diversifying selection contribute to the evolution of the RPP8 gene in Arabidopsis thaliana
A/Reference number: Z24999; MUID:99030193; PMID:9811794
A/Accession: T48899
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-908 <MCD>
A/Cross-references: EMBL:AF089711; NID:G3901293; PIDN:AAK8631.1; PID:G3901294
A/Experimental source: Columbia
C/Genetics:
A/Gene: rpp8
A/Introns: 293/1; 342/1
C/Function:
A/Description: susceptible allele of a gene that promotes resistance to Peronospora parasitica

Query Match      70.0%; Score 35; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
   |||
Db 883 EKLVPQGEDY 892

RESULT 7
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C/Species: tobacco yellow dwarf virus
C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C/Accession: A42452
R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.; Virology 187, 633-642, 1992
A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yell. A/Reference number: A42452; MUID:92188538; PMID:1546458
A/Accession: A42452
A/Molecule type: DNA
A/Residues: 1-102 <MOR>

```


N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs prote
 C;Species: simian immunodeficiency virus SIVcpz
 A;Note: host Pan troglodytes (chimpanzee)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: S03988

R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A;Reference number: S03983; MUID:90259077; PMID:2188136
 A;Accession: S03988
 A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
 A;Residues: 1-124 <HUE>
 A;Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
 C;Genetics:

A;Gene: rev; trs; art
 A;Introns: 27/1
 C;Superfamily: AIDS trans-regulatory splicing protein
 C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 66.0%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGXDXS 11
 |||||

DB 107 EVVPPAGNYS 116
 |||||

RESULT 15

AG1272

thiol peroxidases homolog lmo1583 [imported] - Listeria monocytogenes (strain EGD-e)
 C;Species: Listeria monocytogenes
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C;Accession: AG1272

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloecke
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
 Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
 A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AG1272

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-165 <GLA>

A;Cross-references: GB:NC_003210; PIDN:CAC99661.1; PID:gl6411012; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo1583

C;Superfamily: thioredoxin peroxidase

Query Match 66.0%; Score 33; DB 2; Length 165;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGXDX 10
 |||||

DB 144 EVVPEGSDH 152
 |||||

Search completed: June 3, 2004, 12:00:03
 Job time : 10 secs

QY 2 EVVVPXGXDX 10
 |||||

DB 194 EVTPEGSDY 202
 |||||

RESULT 12

H87660

peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: H87660

R;Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: H87660

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-433 <STO>

A;Cross-references: GB:AE005673; NID:gl3425020; PIDN:AAK25284.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC3322

Query Match 68.0%; Score 34; DB 2; Length 433;
 Best Local Similarity 54.5%; Pred. No. 30;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
 |||||

DB 266 EVILPPGFDIS 276
 |||||

RESULT 13

B49132

fat facets (faf) splice form 1 - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 01-Dec-2000

C;Accession: B49132; A49132

R;Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.

Development 116, 985-1000, 1992

A;Title: The fat facets gene is required for Drosophila eye and embryo development.

A;Reference number: A49132; MUID:93202020; PMID:1295747

A;Contents: isogenic st

A;Accession: B49132

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-2747 <FIS>

A;Cross-references: GB:L04959; NID:G157411; PIDN:AAF01345.1; PID:G6013474

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:127839)

A;Accession: A49132

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-2704, 'VT', '2707', 'ANNV' <FI2>

A;Cross-references: GB:L04958; NID:G157410; PIDN:AAF01346.1; PID:G6013475

A;Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:129029)

C;Keywords: alternative splicing

Query Match 68.0%; Score 34; DB 2; Length 2747;

Best Local Similarity 54.5%; Pred. No. 2.2e+02;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
 |||||

DB 1394 EVIPEGQDPS 1404
 |||||

RESULT 14

VKLJSI

trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds

(without alignments)

117.693 Million cell updates/sec

Title: US-09-909-164-45

Perfect score: 50

Sequence: 1 BEVVPXGXDXS 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35	70.0	908	1 R8L4 ARATH	Q9fjK8 arabidopsis
2	35	70.0	908	1 RPP8 ARATH	Q8w4j9 arabidopsis
3	35	70.0	910	1 RPP8 ARATH	P59584 arabidopsis
4	35	70.0	1058	1 CARB_FUSNN	Q9rg86 fusobacteri
5	34	68.0	102	1 Y1IK Y1DVA	P1619 tobacco yel
6	34	68.0	394	1 HNEA YIBCH	Q9km33 vibrio chol
7	34	68.0	2778	1 FAF DROME	P55824 drosophila
8	33	66.0	124	1 REV STVZ	P17280 chimpanzee
9	33	66.0	165	1 TPX LISIN	Q92bc5 listeria in
10	33	66.0	165	1 TPX LISMO	Q8v6u8 listeria mo
11	33	66.0	327	1 YK14 CAEEL	P4338 caenorhabdi
12	33	66.0	421	1 ACDM RAT	P08503 rattus norv
13	33	66.0	563	1 SYR STRP3	Q8k512 streptococ
14	33	66.0	563	1 SYR STRP8	Q8n222 streptococ
15	33	66.0	563	1 SYR STRPY	Q99x15 streptococ
16	32	64.0	154	1 PLAS ORYSA	P20423 oryza sativ
17	32	64.0	155	1 PLAS HORVU	P08248 hordeum vul
18	32	64.0	168	1 PLAT POPNI	P1970 populus nig
19	32	64.0	196	1 YJAG SALTY	Q91912 salmonella
20	32	64.0	302	1 DDL GLOVI	Q7nmw1 gloeobacter
21	32	64.0	336	1 SUGT MOUSE	Q9cx34 mus musculu
22	32	64.0	427	1 TOLB HAEIN	P44677 haemophilus
23	32	64.0	1499	1 A10A HUMAN	O60312 homo sapien
24	32	64.0	3174	1 CHAC HUMAN	Q96r17 homo sapien
25	31	62.0	233	1 H1S9 THEWA	Q9wzr1 thermotoga
26	31	62.0	319	1 YHAI CRYPA	P10941 cryptospectr
27	31	62.0	421	1 ACDM HUMAN	P11310 homo sapien
28	31	62.0	421	1 ACDM PIG	P41367 sus scrofa
29	31	62.0	562	1 TR2M PANAY	Q47861 pantoea agg
30	30	60.0	97	1 PLAS DAUCA	P20422 daucus caro
31	30	60.0	121	1 TKNK HUMAN	Q9uhf0 homo sapien
32	30	60.0	165	1 YJ49 ARCFU	Q29330 archaeoglob
33	30	60.0	175	1 HES3 RAT	Q04667 rattus norv

34	30	60.0	203	1 CTC BACSU	P14194 bacillus su
35	30	60.0	231	1 ARAD_ECOLI	P08203 escherichia
36	30	60.0	231	1 ARAD_SALTY	P06190 salmonella
37	30	60.0	232	1 SCOA_HELPJ	Q9zle3 helicobacte
38	30	60.0	232	1 SCOA_HELPY	P56006 helicobacte
39	30	60.0	259	1 OVUH_LYMST	P06308 lymnaea sta
40	30	60.0	421	1 AMP2_YEAST	P38174 saccharomyc
41	30	60.0	421	1 ECB2_HALEL	O52250 halomonas e
42	30	60.0	423	1 ECB1_HALEL	Q9zeu7 halomonas e
43	30	60.0	430	1 FOLC_BACSU	Q05865 bacillus su
44	30	60.0	457	1 Z185 HUMAN	O15231 homo sapien
45	30	60.0	469	1 LETI_KLULA	P53998 kluyveromyc

ALIGNMENTS

RESULT 1
R8L4 ARATH
ID R8L4 ARATH STANDARD; PRT; 908 AA.
AC Q9RUK8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR ATSG48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Seki M., Tida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RL "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Potential disease resistance protein.
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (By similarity).
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrrs.ucdavis.edu".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB015468; BAB10695.1; -
CC EMBL; AK117163; BAC41841.1; -
CC InterPro; IPR000767; Disease_resist.
CC InterPro; IPR001611; LRR.

```

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brookes S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjail M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T.T., Kawai J.,
RA Kaniya A., Meyers C., Nakajima M., Naruseka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender B.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RT Science 302:842-846 (2003).
RN [5]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RN STRAIN:cv. Columbia;
RC Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Naruseka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RA "Arabidopsis thaliana full-length cDNA.";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN INTERACTION WITH TIP.
RN MEDLINE=20496823; PubMedId=1041886;
RA Ren T., Qu F., Morris T.J.;
RT "HRT gene function requires interaction between a NAC protein and
RT viral capsid protein to confer resistance to turnip crinkle virus.";
RT Plant Cell 12:1917-1926 (2000).
CC -|- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. The interaction
CC with TIP (TGV-interacting protein) may be essential for the
CC recognition of the avirulence proteins, and the triggering of the
CC defense response
CC -|- SUBUNIT: Interacts with the NAC protein TIP.
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8WAJ9-1; SequencesDisplayed;
CC Name=2;
CC IsoId=Q8WAJ9-2; Sequences=VSP 007171, VSP 007172;
CC Note=Has been shown to exist only in cv "Columbia so far;
CC -|- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition.
CC -|- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
CC cv. Columbia are probably due to an unequal crossing-over between
CC the highly related RPP8 and RPP8A genes present in cv. Landsberg
CC erecta. Such variations probably modify the specificity of
CC pathogen recognition.
CC -|- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
CC specifically recognizes the Emco5 avirulence protein from
CC Parospora parasitica, while it is not the case in cv. Di-17,
CC where it confers resistance to Turnip Crinkle Virus upon
CC recognition of the viral capsid protein.
CC -|- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -|- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -|- SIMILARITY: Contains 1 NB-ARC domain.
CC -|- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homeologs;
CC WWW="http://niblrrs.ucdavis.edu".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF089710; AAC83165.1; --

```


Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVVPXGXDY 10
 Db 885 EKLVPGGEDY 894

RESULT 4
 CARB_FUSNN STANDARD; PRT; 1058 AA.
 ID CARB_FUSNN STANDARD; PRT; 1058 AA.
 AC QBRG86;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 DE Carbamoyl-phosphate synthase ammonia chain).
 GN CARB OR FN0422.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OC Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]

SEQUENCE FROM N.A.
 RP STRAIN=ATCC 25586;
 RC MEDLINE=21886394; PubMed=11889109;
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fostein M., Rykides N., Overbeek R.;
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RT J. Bacteriol. 184:2005-2018(2002).
 RL -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 promotes the hydrolysis of glutamine to ammonia, which is used by
 the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 similarity).

-1- SIMILARITY: Belongs to the carb family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AB010554; AAJ94625.1; ALT_INIT.
 HAWAP; MF_01210; -; 1.
 InterPro; IPR006275; CarA_L_glu.
 InterPro; IPR005483; CPase_L.
 InterPro; IPR005479; CPase_L_D2.
 InterPro; IPR005480; CPase_L_D3.
 InterPro; IPR005481; CPase_L_N.
 InterPro; IPR004362; MGS_like.
 Pfam; PF00289; CPase_L_Chain; 2.
 Pfam; PF02786; CPase_L_D2; 2.
 Pfam; PF02787; CPase_L_D3; 2.
 Pfam; PF02142; MGS_1_D3; 1.
 PRINTS; PR00098; CPASE.
 TIGR01369; CPaseII_lrg; 1.
 PROSITE; PS00866; CPASE_1; 2.
 PROSITE; PS00867; CPASE_2; 2.
 Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401
 FT DOMAIN 402 546
 FT DOMAIN 547 529
 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.

FT DOMAIN 930 1058
 REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210
 FT NP_BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1B39F CRC64;

Query Match 70.0%; Score 35; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 22;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVVPXGXDYS 11
 Db 190 EIVPGLNYS 199

RESULT 5
 Y11K_TYDVA STANDARD; PRT; 102 AA.
 ID Y11K_TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Y1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 OX NCBI_TaxID=31599;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92188638; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RA "The nucleotide sequence of the infectious cloned DNA component of
 tobacco yellow dwarf virus reveals features of geminiviruses
 infecting monocotyledonous plants.";
 RL Virology 187:633-642(1992).
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M81103; AAA47947.1; -.
 PIR; A42452; A42452.
 InterPro; IPR002621; Gemini_mov.
 Pfam; PF01708; Gemini_mov; 1.
 Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVVPXGXDYS 11
 Db 7 QVVPESGINS 16

RESULT 6
 HMPA_VIBCH STANDARD; PRT; 394 AA.
 ID HMPA_VIBCH STANDARD; PRT; 394 AA.
 AC Q9KMY3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

DE Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin).
GN HMP OR VCA0183.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=EL Tor N16961 / Serotype O1;
RC MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:1477-483(2000).
CC -!- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
CC OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
CC PAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION.
CC -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO-DOMAIN
CC FLAVOHEMOPROTEINS SUBFAMILY.
CC -!- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
CC OXIDOREDUCTASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004358; NAP96096.1; -.
DR F1R; F82491; F82491.
DR HSP; F39662; 1COX.
DR TIGR; VCA0183; -.
DR InterPro; IPR008333; FAD binding_6.
DR InterPro; IPR001709; FPN_cyt_redctse.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR001433; Oxred_FAD/NAD(P).
DR InterPro; IPR001221; Phe_hydroxylase.
DR Pfam; PF00970; FAD_binding_6; 1.
DR Pfam; PF00042; Globin; 1.
DR Pfam; PF00175; NAD_binding_1; 1.
DR PRINTS; PR00371; FENCR.
DR PRINTS; PR00410; PHEHYDRLASE.
DR PROSITE; PS01033; GLOBIN; 1.
KW Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport;
KW Oxygen transport; Transport; Complete proteome.
FT DOMAIN 1 136
FT METAL 53 53
FT METAL 85 85
FT METAL 85 85
FT NP_BIND 268 273 NADP (KMOSE PART) (BY SIMILARITY).
FT SEQUENCE 394 AA; 44191 MW; DDA3490FAE28823A CRC64;
SQ
Query Match 68.0%; Score 34; DB 1; Length 394;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVPGXDY 10
DB 194 EVTPEGSY 202
RESULT 7
FAF_DROME STANDARD; PRT; 2778 AA.
ID _FAF_DROME

```

P55824; Q9V9T6; Q9Y0Z7;
01-NOV-1997 (Rel. 35, Created)
28-FEB-2003 (Rel. 43, Last annotation update)
15-MAR-2004 (Rel. 43, Last annotation update)
Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
(Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
FAF) (Deubiquitinating enzyme FAF) (Pat facets protein).
GN FAF OR BCDNA.LD22582 OR CG1945
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
RP TISSUE=Eye, imaginal disk;
RC MEDLINE=33202020; PubMed=1295747;
RA Fischer-Vize J.A., Rubin G.M., Lehmann R.;
RT "The fat facets gene is required for Drosophila eye and embryo
RT development.";
RL Development 116:985-1000(1992).
RN [2]_SEQUENCE FROM N.A.
RP STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.S., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]_REVIEWS, AND ALTERNATIVE SPLICING.
RP MEDLINE=22426059; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;

RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]

RP SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).

RC STRAIN=Berkeley;

RX MEDLINE=20196012; PubMed=10731138;

RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,

RA Stapleton M., Harvey D.A.;

RL "A Drosophila complementary DNA resource."

CC Science 287:2222-2224(2000).

CC -!- FUNCTION: Required for eye and embryo development, and plays a
CC role in compound eye assembly and oogenesis respectively. In the
CC larval eye disks, cells outside the assembling facets require this
CC protein for short-range cell interactions that prevent the mystery
CC calls from becoming photoreceptors. It is also required for
CC nuclear migration and cellularization in early embryogenesis and
CC could play a role in pole cell determination, development or
CC function.

CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.

CC -!- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=3;

CC Comment=Experimental confirmation may be lacking for some
CC isoforms;

CC Name=1;

CC IsoId=P55824-1; Sequence=Displayed;

CC Name=2;

CC IsoId=P55824-2; Sequence=VSP_005270;

CC Name=3;

CC IsoId=P55824-3; Sequence=VSP_005269;

CC -!- TISSUE SPECIFICITY: Eye disks and ovaries.

CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.

CC -!- SIMILARITY: Belongs to peptidase family C19.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L04959; AAF01345.1; -

DR EMBL; L04958; AAF01346.1; -

DR EMBL; L04960; AAF01347.1; -

DR EMBL; L04960; AAF01348.1; -

DR EMBL; AE003779; AAF57198.1; -

DR EMBL; AE003779; AAF57198.1; -

DR EMBL; AF145677; AAD38652.1; -

DR MEROPS; C19.007; -

DR FlyBase; FBgn0005632; faf.

DR GO; GO:0005737; Cytoplasm; IDA.

DR GO; GO:0007349; P:cellularization; IMP.

DR GO; GO:0009795; P:embryonic morphogenesis; IMP.

DR GO; GO:0007456; P:eye morphogenesis (sensu Drosophila); IMP.

DR GO; GO:0008583; P:myotery cell fate differentiation (sensu Dr. .; IMP.

DR GO; GO:0007097; P:nuclear migration; IMP.

DR GO; GO:0036579; P:protein deubiquitination; IDA.

DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IGI.

DR InterPro; IPR001394; Peptidase_C19.

DR Pfam; PF00443; UCH; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS00973; UCH_2_2; 1.

DR PROSITE; PS0235; UCH_2_3; 1.

KW Ubl conjugation pathway; Hydrolase; Thiol protease;

KW Developmental protein; Vision; Alternative splicing.

FT ACT SITE 1677 1677 By SIMILARITY.

FT ACT SITE 1978 1978 By SIMILARITY.

FT ACT SITE 1986 1986 By SIMILARITY.

FT VARSPLIC 2705 2778 KCRVILKLVESKDEEDATTATTAATTEVTTSPATAIATA

FT ATLEPAGSELTTWKNLIISQENPOAKSLQ -> VTRA

FT NNV (in isoform 3).

FT

FT VARSPLIC 2742 2778 /FTId=VSP_005269.
FT IATAATLEPAGSELTTWKNLIISQENPOAKSLQ ->
FT SORQOL (in isoform 2).
FT /FTId=VSP_005270.

FT CONFLICT 234 234 E -> D (IN REF. 1).

FT CONFLICT 2725 2725 T -> S (IN REF. 1; AAF01345).

FT SEQUENCE 2778 AA; 311139 MW; FFB90438BA53A02B CRC64;

Query Match 68.0%; Score 34; DB 1; Length 2778;

Best Local Similarity 54.5%; Pred. No. 97;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGDYS 11

Db 1394 EVIVPDQDFS 1404

RESULT 8

REV_SIVCZ

ID -REV_SIVCZ STANDARD; PRT; 124 AA.

AC P17280;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE REV protein (Anti-repression transactivator protein) (ART/TRS).

GN REV.

OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11723;

RP SEQUENCE FROM N.A.

RX MEDLINE=90259077; PubMed=2188136;

RA Huet T., Cheynier R., Meyerhans A., Roelants G., Main-Hobson S.;

RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";

RL Nature 345:356-359(1990).

CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE

CC -!- NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.

CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.

CC -!- PTM: Phosphoprotein whose state of phosphorylation is mediated by

CC a specific serine kinase activity present in the nucleus.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X52154; CAA36405.1; -

DR PIR; S09588; VKLJST.

DR HIV; X52154; REVSCPZ.

DR InterPro; IPR000625; REV_protein.

DR Pfam; PF00424; REV; 1.

DR Transcription regulation; AIDS; Phosphorylation; Nuclear protein.

SK SEQUENCE 124 AA; 13701 MW; F5877DIBDP65A7B2 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 124;

Best Local Similarity 60.0%; Pred. No. 6.2;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVVPXGDYS 11

Db 107 ETVFAGNYS 116

RESULT 9

TPX_LISIN

ID -TPX_LISIN STANDARD; PRT; 165 AA.

AC Q92BC5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable thiol peroxidase (EC 1.11.1.1.-).

GN TPX OR LIN1625.

OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1642;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of *Listeria species*."

RL Science 294:849-852(2001).

CC -!- FUNCTION: Has antioxidant activity. Could remove peroxidases or

CC -!- H(2)O(2) (By similarity).

CC -!- SIMILARITY: Belongs to the ahpc/tsa family. Tpx subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AL596169; CAC96856.1; -.

DR PIR; AG1272; AG1272.

DR ListList; LMO01583; -.

DR HAMAP; MF 00269; -; 1.

DR InterPro; IPR000866; Ahpc-TSA.

DR InterPro; IPR002065; TPX.

DR Pfam; PF00578; Ahpc-TSA; 1.

DR PROSITE; PS01265; TPX; 1.

KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.

SQ SEQUENCE 165 AA; 18162 MW; 77705B7CD8BC8F4D CRC64;

Query Match 66.0%; Score 33; DB 1; Length 165;

Best Local Similarity 66.7%; Pred. No. 8.4;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXGXDY 10

DB 144 EVVPEGS DH 152

RESULT 10

TPX LISMO STANDARD; PRT; 165 AA.

AC Q8YFUS.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable thiol peroxidase (EC 1.11.1.-).

GN TPX OR LMO01583.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1639;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EGD-e / Serovar 1/2a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of *Listeria species*."

RL Science 294:849-852(2001).

CC -!- FUNCTION: Has antioxidant activity. Could remove peroxidases or

CC -!- H(2)O(2) (By similarity).

CC -!- SIMILARITY: Belongs to the ahpc/tsa family. Tpx subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AL591979; CAC99661.1; -.

DR PIR; AG1272; AG1272.

DR ListList; LMO01583; -.

DR HAMAP; MF 00269; -; 1.

DR InterPro; IPR000866; Ahpc-TSA.

DR InterPro; IPR002065; TPX.

DR Pfam; PF00578; Ahpc-TSA; 1.

DR PROSITE; PS01265; TPX; 1.

KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.

SQ SEQUENCE 165 AA; 18133 MW; 77705B7CC46D424D CRC64;

Query Match 66.0%; Score 33; DB 1; Length 165;

Best Local Similarity 66.7%; Pred. No. 8.4;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXGXDY 10

DB 144 EVVPEGS DH 152

RESULT 11

YK14 CAEEL STANDARD; PRT; 327 AA.

AC P34338;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C15H7.4 in chromosome III.

GN C15H7.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,

RA Paterson P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,

RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

RA Waterston R., Watson A., Weinstein L., Wilkinson-Spratt J.,

RA Woldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."

RL Nature 368:32-38(1994).


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 222173; CAA80126.1; -
DR PIR; S40753; S40753.
DR WormPep; CL5H7.4; CE000082.
KW Hypothetical protein.
SQ SEQUENCE 327 AA; 35366 MW; 716BC2BDD2E9607E CRC64;

Query Match 66.0%; Score 33; DB 1; Length 327;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
DB 175 KEVVPNGGDKS 185

RESULT 12
ACDM_RAT ACDM_RAT STANDARD; PRT; 421 AA.
AC P08503;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acyl-CoA dehydrogenase, medium-chain specific, mitochondrial precursor
DE (EC 1.3.99.3) (MCAD).
GN ACADM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=87280028; PubMed=3611054;
RA Matsubara Y., Kraus J.P., Ozasa H., Glassberg R., Finocchiaro G.,
RA Ikeda Y., Mole J., Rosenberg L.E., Tanaka K.;
RT "Molecular cloning and nucleotide sequence of cDNA encoding the
RT entire precursor of rat liver medium chain acyl coenzyme A
RT dehydrogenase."
RL J. Biol. Chem. 262:10104-10108(1987).
CC 16.
CC -!- FUNCTION: This enzyme is specific for acyl chain lengths of 4 to
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + ETF = 2,3-dehydroacyl-CoA + reduced
CC ETF.
CC -!- COFACTOR: FAD.
CC -!- PATHWAY: Mitochondrial fatty acid beta-oxidation system; first
CC step.
CC -!- SUBUNIT: Homotrimer.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- MISCELLANEOUS: A number of straight-chain acyl-CoA dehydrogenases
CC of different substrate specificities are present in mammalian
CC tissues.
CC -!- SIMILARITY: Belongs to the acyl-CoA dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02791; AAA40670.1; -
DR PIR; A28436; DERTCM.
DR HSSP; P11310; 1EGD.

```

```

DR InterPro; IPR006089; Acyl-CoA dh.
DR InterPro; IPR006090; Acyl-CoA dh. C.
DR InterPro; IPR006091; Acyl-CoA dh. M.
DR InterPro; IPR006092; Acyl-CoA dh. N.
DR Pfam; PF00441; Acyl-CoA dh. 1.
DR Pfam; PF02770; Acyl-CoA dh. M. 1.
DR Pfam; PF02771; Acyl-CoA dh. N. 1.
DR PROSITE; PS00072; ACYL COA DH 1; 1.
DR PROSITE; PS00073; ACYL COA DH 2; 1.
KW Oxidoreductase; Flavoprotein; FAD; Fatty acid metabolism;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 25 MITOCHONDRION.
FT CHAIN 26 421 ACYL-CoA DEHYDROGENASE, MEDIUM-CHAIN
FT ACT_SITE 193 193 SPECIFIC.
FT ACT_SITE 401 401 FORMS A HYDROGEN-BOND WITH THE FLAVIN
FT ACT_SITE 401 401 N(5) OF THE FAD COFACTOR (BY SIMILARITY).
SQ SEQUENCE 421 AA; 46555 MW; 2CF076F8C919BDE8 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 421;
Best Local Similarity 50.8%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
DB 58 EEIIPVAPDY 67

RESULT 13
SYR_STRP3 SYR_STRP3 STANDARD; PRT; 563 AA.
AC Q8K5J2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARGRS OR SPYM3_1809 OR SP51807.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22693278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Iasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1043-1055 (2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (by similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF014171; AAM80416.1; -
EMBL; AF005146; BAC64902.1; -
HAMAP; MF_00123; -; 1.
InterPro; IPR001278; Arg_trna-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; trna-synt_ld_C.
InterPro; IPR001412; trna-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; trna-synt_ld; 1.
Pfam; PF05746; trna-synt_ld_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMs; TIGR00456; args; 1.
PROSITE; PS00178; AA TRNA_LIGASE_I; FALSE NEG.
Aminoacyl-trna synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 121 131 "HIGH" REGION.
SQ SEQUENCE 563 AA; 63134 MW; B0B7DEC31A9DCF63 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
Best Local Similarity 45.5%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
|:|:|:|:
Db 94 EQVITAGSDYA 104

RESULT 14
SYN_STRP8
ID SYN_STRP8 STANDARD; PRT; 563 AA.
AC Q8NZ22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-trna synthetase (EC 6.1.1.19) (Arginine--trna ligase) (ArgRS).
GN ARG5 OR SPYM18.2183.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S., Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F., Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q., Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + trna(Arg) = AMP + diphosphate + L-arginyl-trna(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to class-I aminoacyl-trna synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF010119; RAL95627.1; -
HAMAP; MF_00123; -; 1.

InterPro; IPR001278; Arg_trna-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; trna-synt_ld_C.
InterPro; IPR001412; trna-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; trna-synt_ld; 1.
Pfam; PF05746; trna-synt_ld_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMs; TIGR00456; args; 1.
PROSITE; PS00178; AA TRNA_LIGASE_I; FALSE NEG.
Aminoacyl-trna synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 121 131 "HIGH" REGION.
SQ SEQUENCE 563 AA; 63134 MW; FECAPI76A69D8B5B CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
Best Local Similarity 45.5%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
|:|:|:|:
Db 94 EQVITAGSDYA 104

RESULT 15
SYN_STRPY
ID SYN_STRPY STANDARD; PRT; 563 AA.
AC Q99XL5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-trna synthetase (EC 6.1.1.19) (Arginine--trna ligase) (ArgRS).
GN ARG5 OR SPY2151.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvoarov A.N., Kenyon S., Lai H.S., Lin S.F., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + trna(Arg) = AMP + diphosphate + L-arginyl-trna(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-trna synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE006633; AAK34788.1; -
HAMAP; MF_00123; -; 1.
InterPro; IPR001278; Arg_trna-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; trna-synt_ld_C.
InterPro; IPR001412; trna-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; trna-synt_ld; 1.
Pfam; PF05746; trna-synt_ld_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMs; TIGR00456; args; 1.

DR PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 FT SITE 121 131
 SQ SEQUENCE 563 AA; 63120 MW; E0F2CAC28D03B613 CRC64;
 "HIGH" REGION.
 Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 30;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVPXGXDYS 11
 Db 94 EQVITAGSDYA 104

Search completed: June 3, 2004, 11:49:56
 Job time : 5.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds

(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-45

Perfect score: 50

Sequence: 1 BEVVPXGXDS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organella.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	37	74.0	149	16	Q82ZB4	Q82ZB4 enterococu
2	26	72.0	298	2	O52367	O52367 rhizobium t
3	36	72.0	319	16	Q8SV7	Q8SV7 oceanobacil
4	36	72.0	363	17	O30260	O30260 archaeeoglob
5	36	72.0	595	10	Q94637	Q94637 andrographi
6	36	72.0	1044	16	Q8DIH0	Q8DIH0 synchococ
7	35	70.0	78	6	Q9XST4	Q9XST4 canis famil
8	35	70.0	175	6	Q81033	Q81033 bos taurus
9	35	70.0	215	6	Q81031	Q81031 bos taurus
10	35	70.0	217	4	O00404	O00404 homo sapien
11	35	70.0	281	6	Q867A5	Q867A5 tragulus ja
12	35	70.0	290	16	Q8U7J0	Q8U7J0 agrobacteri
13	35	70.0	297	6	Q8HYJ9	Q8HYJ9 bos taurus
14	35	70.0	299	4	Q8UEE9	Q8UEE9 homo sapien
15	35	70.0	587	16	Q9JZP8	Q9JZP8 neisseria m
16	35	70.0	592	5	Q9VI90	Q9VI90 drosophila

17	35	70.0	905	10	Q84V54	Q84V54 arabidopsis
18	34	68.0	156	3	Q12479	Q12479 saccharomyc
19	34	68.0	175	10	Q8V188	Q8V188 arabidopsis
20	34	68.0	193	9	Q859A6	Q859A6 bacterioph
21	34	68.0	247	16	Q92U66	Q92U66 rhizobium m
22	34	68.0	257	10	Q9C6J0	Q9C6J0 arabidopsis
23	34	68.0	283	16	Q8NR16	Q8NR16 corynebacte
24	34	68.0	322	16	Q82DT4	Q82DT4 streptomyce
25	34	68.0	433	16	Q9A382	Q9A382 caulobacter
26	34	68.0	1442	17	Q96VH5	Q96VH5 sulfolobus
27	33	66.0	143	17	Q8FX62	Q8FX62 methanopyru
28	33	66.0	196	16	Q82AQ9	Q82AQ9 versinia pe
29	33	66.0	210	10	Q85890	Q85890 cyclostell
30	33	66.0	222	16	Q8E5L9	Q8E5L9 streptococ
31	33	66.0	222	16	Q8DZM9	Q8DZM9 streptococ
32	33	66.0	225	10	Q40129	Q40129 lycopersico
33	33	66.0	253	16	Q8XPA8	Q8XPA8 clostridium
34	33	66.0	283	16	Q8FPV9	Q8FPV9 corynebacte
35	33	66.0	308	16	Q83NN9	Q83NN9 tropheryma
36	33	66.0	308	16	Q83G59	Q83G59 tropheryma
37	33	66.0	368	2	Q8KL43	Q8KL43 rhizobium e
38	33	66.0	415	16	Q7VGY0	Q7VGY0 helicobacte
39	33	66.0	440	17	Q9VFI3	Q9VFI3 aeropyrum p
40	33	66.0	471	11	Q8R126	Q8R126 mus musculu
41	33	66.0	484	11	Q8VD18	Q8VD18 mus musculu
42	33	66.0	484	11	Q8BFX4	Q8BFX4 mus musculu
43	33	66.0	484	11	Q8BK35	Q8BK35 mus musculu
44	33	66.0	517	16	Q8XZL5	Q8XZL5 raistonia s
45	33	66.0	563	13	Q9DDJ4	Q9DDJ4 halichoeres

ALIGNMENTS

RESULT 1

Q82ZB4 PRELIMINARY; PRT; 149 AA.
AC Q82ZB4;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Large conductance mechanosensitive channel protein.
GN MSL OR EF3152.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]_TaxID=1351;
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RK MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016956; AA082828.1; --
DR TIGR; EF3152; --
DR GO; GO:0004602; C:membrane; IEA.
DR GO; GO:0005215; F:ion channel activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001185; MS_channel.
DR Pfam; PF01741; MscL; 1.
DR PRINTS; PR01264; MECHCHANNEL.
DR ProDom; PD007253; MS_channel; 1.
DR TIGRFAMs; TIGR00220; mscL; 1.
KW Complete proteome.
SQ SEQUENCE 149 AA; 16127 MW; 555799BF1E47D34E CRC64;

Query Match

74.0%; Score 37; DB 16; Length 149;

```
Best Local Similarity 70.0%; Pred. No. 5.2;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVPXGXDY 10
Db 113 EVVPTSEDY 122

RESULT 2
O52367 PRELIMINARY; PRT; 298 AA.
ID O52367
AC O52367
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aryl-alcohol dehydrogenase homolog (Fragment).
DE XYLBI.
GN Rhizobium tropici.
OS Rhizobium tropici.
OG Plasmid pRtrCFN299a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CFN299;
RA Rosenbluth M., Hynes M.F., Martinez-Romero E.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL; AF036920; AAC04779.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity; zinc-dependent; IEA.
DR GO; GO:0004649; F:alcohol dehydrogenase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; NAD_BS.
DR InterPro; IPR00205; ADH_zinc_N.
DR Pfam; PF00107; ADH_zinc_N.1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Metal-binding; Oxidoreductase; Zinc; Plasmid.
FT NON TER 298
SQ SEQUENCE 298 AA; 31092 MW; 49B2F8117C3AB87 CRC64;

Query Match 72.0%; Score 36; DB 2; Length 298;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGXDYS 11
Db 250 EIIPEGADFS 259

RESULT 3
Q8ESV7 PRELIMINARY; PRT; 319 AA.
ID Q8ESV7
AC Q8ESV7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical conserved protein.
GN OB0509.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HT833 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12253376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
```

```
RL Nucleic Acids Res. 30:3927-3935 (2002).
DR EMBL; AP004594; BAC12465.1; -.
DR InterPro; IPR001279; Blackmase-like.
DR Pfam; PF00753; lactamase_B; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDA4BF13E79E37 CRC64;

Query Match 72.0%; Score 36; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVPXGXDY 10
Db 189 EQLVPHGIDY 198

RESULT 4
O30260 PRELIMINARY; PRT; 363 AA.
ID O30260
AC O30260
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein AF2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370 (1997).
DR EMBL; AS001109; AAB91255.1; -.
DR PIR; D69551; D69551.
DR TIGR; AF2411; -.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0E9762AE788F4803 CRC64;

Query Match 72.0%; Score 36; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVPXGXDYS 11
Db 120 ENVPYGDIFS 130

RESULT 5
Q946J7 PRELIMINARY; PRT; 595 AA.
ID Q946J7
AC Q946J7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (EC 1.1.1.34).
GN HMGRI.
```

OS Andrographis paniculata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Lamiaceae; Acanthaceae; Acanthoideae; Ruellieae;
 OC Andrographis; Andrographis.
 OX NCBI_TaxID=175694;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krishnan S., Banerjee N.S.;
 RT "3-hydroxy-3-methylglutaryl coenzyme A reductase gene from
 RT Andrographis paniculata."
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF389879; AL28015.2; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0004420; F:hydroxymethylglutaryl-CoA reductase (NADPH) . . .; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0003058; F:biosynthesis; IEA.
 DR GO; GO:0006629; P:lipid metabolism; IEA.
 DR InterPro; IPR002022; HMG-CoA red.
 DR InterPro; IPR009023; HMG-CoA_NAD_bind.
 DR InterPro; IPR004554; HMG-CoA_NADP.
 DR InterPro; IPR009029; HMG-CoA_sub_bind.
 DR Pfam; PF00368; HMG-CoA_red; 1.
 DR PRINTS; PR00071; HMGCOARCTASE.
 DR TIGRFAMs; TIGR00533; HMG COA R_NADP; 1.
 DR PROSITE; PS00066; HMG COA REDUCTASE_1; 1.
 DR PROSITE; PS00318; HMG COA REDUCTASE_2; 1.
 DR PROSITE; PS01192; HMG COA REDUCTASE_3; 1.
 DR PROSITE; PS00665; HMG COA REDUCTASE_4; 1.
 DR Oxidoreductase.
 KW Oxidoreductase.
 SQ SEQUENCE 595 AA; 63268 MW; 19A3EA572F67AB2E CRC64;
 Query Match 72.0%; Score 36; DB 10; Length 595;
 Best Local Similarity 70.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVVPXGXDYS 11
 DB 15 EVAPPCHDYS 24
 RESULT 6
 Q8DIHO PRELIMINARY; PRT; 1044 AA.
 ID Q8DIHO;
 AC Q8DIHO;
 DT 01-NAR-2003 (TRENBLrel. 23, Created)
 DT 01-NAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLL1618.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1."
 RT DNA Res. 9:123-130(2002).
 RL EMBL; AF005374; BAC09170.1; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001036; Acflvin_res.
 DR InterPro; IPR004764; HSL.
 DR Pfam; PF00873; ACR_tran; 1.
 DR PRINTS; PR00702; ACRIFLAVINRP.
 DR TIGRFAMs; TIGR00915; 2A0602; 1.

KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
 Query Match 72.0%; Score 36; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 79;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGXDYS 11
 DB 843 BEVLPNGIGYS 853
 RESULT 7
 Q9XST4 PRELIMINARY; PRT; 78 AA.
 ID Q9XST4;
 AC Q9XST4;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE P97 homologous protein (Fragment).
 GN P97;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thyroid;
 RX MEDLINE=20422104; PubMed=10964405;
 RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
 RA Christophe D.;
 RT "A method for the large-scale cloning of nuclear proteins and nuclear
 RT targeting sequences on a functional basis."
 RL Anal. Biochem. 284:231-239(2000).
 DR EMBL; AJ388531; CAB46833.1; --
 FT NON TER 78
 SQ SEQUENCE 78 AA; 8895 MW; B62486313555FEA1 CRC64;
 Query Match 70.0%; Score 35; DB 6; Length 78;
 Best Local Similarity 54.5%; Pred. No. 6.8;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGXDYS 11
 DB 16 EDYVPSGGEYS 26
 RESULT 8
 Q8I033 PRELIMINARY; PRT; 175 AA.
 ID Q8I033;
 AC Q8I033;
 DT 01-NAR-2003 (TRENBLrel. 23, Created)
 DT 01-NAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Human-type bcl2 protein.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jersey;
 RA Iwashita S., Itoh T.;
 RT "A LINE-mediated gene diversity."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB081003; BAC11952.1; --
 SQ SEQUENCE 175 AA; 19529 MW; CE7283CB9893BB6 CRC64;
 Query Match 70.0%; Score 35; DB 6; Length 175;
 Best Local Similarity 54.5%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 BEVVPXGXDYS 11
Db 16 EDVPSGGEYS 26

RESULT 9
Q81031
ID Q81031 PRELIMINARY; PRT; 215 AA.
AC Q81031;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE H-type bent protein (Fragment).
DE H-TYPE BCNT.
GN Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RZPD #750;
RA Iwashita S., Itoh T., Sezaki M., Oshima K., Hashimoto E.,
RA Kitagawawa Y., Takahashi T., Masui T., Hashimoto K.;
RT "A LINE-mediated gene diversity";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081095; BAC1593.1; -.
FT NON_TER 215 215
SQ SEQUENCE 215 AA; 23822 MW; 7287C89898D05C8 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 215;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
Db 16 EDVPSGGEYS 26

RESULT 10
O00404
ID O00404 PRELIMINARY; PRT; 217 AA.
AC O00404;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P97 homologous protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97160586; PubMed=9006920;
RA Nobukuni T., Kobayashi M., Omori A., Ichinose S., Iwanaga T.,
RA Takahashi I., Takahashi K., Hattori S., Kaibuchi K., Miyata Y.;
RT "An Alu-linked repetitive sequence corresponding to 280 amino acids is
RT expressed in a novel bovine protein, but not in its human homologue.";
RL J. Biol. Chem. 272:2801-2807(1997).
DR EMBL; D85939; BAA20069.1; -.
FT NON_TER 217 217
SQ SEQUENCE 217 AA; 24061 MW; E404BA2E35497828 CRC64;

Query Match 70.0%; Score 35; DB 4; Length 217;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
Db 16 EDVPSGGEYS 26

RESULT 11

```

```

Q867A5
ID Q867A5 PRELIMINARY; PRT; 281 AA.
AC Q867A5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Human-type Bcnt (Fragment).
DE Tragus javanicus (lesser Malay chevrotain).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Tragulina;
OC Tragulidae; Tragulus.
OX NCBI_TaxID=9849;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Ueno S., Kimura J., Kurohmaru M., Fukuta K., Iwashita S.;
RT "Gene organization of the chevrotain bcnt whose paralogue in
RT ruminantia includes an endonuclease domain of RTE-1 in the protein.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB103377; BAC57061.1; -.
FT NON_TER 281 281
SQ SEQUENCE 281 AA; 31557 MW; 8133A9BAFF7509A6 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 281;
Best Local Similarity 54.5%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
Db 8 EDVPSGGEYS 18

RESULT 12
Q8U7J0
ID Q8U7J0 PRELIMINARY; PRT; 290 AA.
AC Q8U7J0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 6-O-methylguanine-DNA methyltransferase.
GN ADA OR ATU4459 OR AGR L 818
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RN Science 294:2317-2323(2001).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Roumel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";

```

```

RL Science 294:2323-2328(2001).
DR EMBL; AE009374; AAL45253.1; -.
DR EMBL; AE008240; AAK88982.1; -.
DR PIR; AG3104; AG3104.
DR PIR; D98182; D98182.
DR GO; GO:0003622; Cinnaracellar; IEA.
DR GO; GO:0003908; F-methylated-DNA-[protein]-cysteine S-methylt. . .; IEA.
DR GO; GO:0008168; F-methyltransferase activity; IEA.
DR GO; GO:0003700; F-transcription factor activity; IEA.
DR GO; GO:0016740; F-transferase activity; IEA.
DR GO; GO:0006281; F-DNA repair; IEA.
DR GO; GO:0006355; P-DNA repair; IEA.
DR InterPro; IPR000005; HTHARAC.
DR InterPro; IPR001497; Methyltransf_1.
DR Pfam; PF001650; HTH_Arac; 2.
DR SMART; SM00342; HTH_ARAC; 1.
DR TIGRfam; TIGR00599; OG1.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 290 AA; 31587 MW; B626592EP519977F CRC64;

Query Match 70.0%; Score 35; DB 16; Length 290;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
DB 9 EDITPGSDY 18

RESULT 13
Q8HXY9 PRELIMINARY; PRT; 297 AA.
AC Q8HXY9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Human-type Bcmt.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Iwashita S., Itoh T.;
RT "A LINE-mediated gene diversity.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081004; BAC1953.1; -.
SQ SEQUENCE 297 AA; 33354 MW; DA9448C8740373C CRC64;

Query Match 70.0%; Score 35; DB 6; Length 297;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 11
DB 16 EDVPSGGGEYS 26

RESULT 14
Q9UEE9 PRELIMINARY; PRT; 299 AA.
AC Q9UEE9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BCNT protein (CRANIOFACIAL development protein 1).
GN BCNT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=9602175;
RA Takahashi I., Nobukuni T., Ohmori H., Kobayashi M., Tanaka S.,
RA Ohshima K., Okada N., Masui T., Hashimoto K., Iwashita S.;
RT "Existence of a bovine LINE repetitive insert that appears in the cDNA
RT of bovine protein BCNT in ruminant, but not in human, genomes.";
RL Gene 211:387-394(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kalline N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB009285; BAA31867.1; -.
DR EMBL; BC000991; AAH00991.1; -.
DR EMBL; BT009819; AAP88821.1; -.
DR Genew; HGNC:1873; CFPD1.
SQ SEQUENCE 299 AA; 33593 MW; F4A9E9288669451A CRC64;

Query Match 70.0%; Score 35; DB 4; Length 299;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 11
DB 16 EDVPSGGGEYS 26

RESULT 15
Q9JZP8 PRELIMINARY; PRT; 587 AA.
AC Q9JZP8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE Succinate dehydrogenase, flavoprotein subunit.
GN NM00950.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MCS8 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Pedersen J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MCS8.";
RL Science 287:1809-1815(2000).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE002446; AAF41356.1; -.
DR PIR; F81138; F81138.
DR HSP; P00363; 1KPF6.
DR TIGR; NM00950; -.
DR GO; GO:0015036; F-disulfide oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR003953; FAD_bind2.
DR InterPro; IPR001327; FAD_Pyr_redox.

```

DR InterPro; IPR003952; FRD/SDH_FAD_BS.
 DR InterPro; IPR001100; Pyr_redox.
 DR InterPro; IPR004112; Succ_DH_flav_C.
 DR Pfam; PF00890; FAD_binding_2; 1.
 DR Pfam; PF02910; succ_DH_flav_C; 1.
 DR PRINTS; PR00368; FADPDR.
 DR PRINTS; PR00411; FNDRTASEI.
 DR PROSITE; PS00504; FRD_SDH_FAD_BINDING; 1.
 KW FAD; Flavoprotein; Oxidoreductase; Complete proteome.
 SQ SEQUENCE 587 AA; 64502 MW; 9581701B08069003 CRC64;

Query Match 70.0%; Score 35; DB 16; Length 587;
 Best Local Similarity 70.0%; Pred No: 68;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
 |||||
 Db 366 EVVVPQGEDY 375

Search completed: June 3, 2004, 11:57:34
 Job time : 29.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model
Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignments)
67.664 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 EEVVPXGDYS 11

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47	94.0	11	5	ABB80548	Hepatitis
2	47	94.0	11	5	ABB80547	Hepatitis
3	47	94.0	11	5	ABB80556	Hepatitis
4	47	94.0	11	5	ABB80557	Hepatitis
5	47	94.0	11	5	ABB80551	Hepatitis
6	46	92.0	11	5	ABB80534	Hepatitis
7	46	92.0	11	5	ABB80561	Hepatitis
8	46	92.0	11	5	ABB80542	Hepatitis
9	46	92.0	11	5	ABB80543	Hepatitis
10	46	92.0	11	5	ABB80524	Hepatitis
11	46	92.0	11	5	ABB80533	Hepatitis
12	46	92.0	11	5	ABB80529	Hepatitis
13	46	92.0	11	5	ABB80528	Hepatitis
14	46	92.0	11	5	ABB80538	Hepatitis
15	46	92.0	11	5	ABB80562	Hepatitis
16	41	82.0	11	5	ABB80549	Hepatitis
17	41	82.0	11	5	ABB80544	Hepatitis
18	41	82.0	11	5	ABB80553	Hepatitis
19	41	82.0	11	5	ABB80552	Hepatitis
20	41	82.0	11	5	ABB80545	Hepatitis
21	40	80.0	11	5	ABB80525	Hepatitis
22	40	80.0	11	5	ABB80521	Hepatitis
23	40	80.0	11	5	ABB80522	Hepatitis
24	40	80.0	11	5	ABB80536	Hepatitis
25	40	80.0	11	5	ABB80566	Hepatitis

26	40	80.0	11	5	ABB80546	Hepatitis
27	40	80.0	11	5	ABB80563	Hepatitis
28	40	80.0	11	5	ABB80565	Hepatitis
29	40	80.0	11	5	ABB80535	Hepatitis
30	40	80.0	11	5	ABB80567	Hepatitis
31	40	80.0	11	5	ABB80540	Hepatitis
32	40	80.0	11	5	ABB80554	Hepatitis
33	40	80.0	11	5	ABB80559	Hepatitis
34	40	80.0	11	5	ABB80526	Hepatitis
35	40	80.0	11	5	ABB80530	Hepatitis
36	40	80.0	11	5	ABB80539	Hepatitis
37	40	80.0	11	5	ABB80550	Hepatitis
38	40	80.0	11	5	ABB80555	Hepatitis
39	40	80.0	11	5	ABB80564	Hepatitis
40	40	80.0	11	5	ABB80568	Hepatitis
41	39	78.0	11	5	ABB80523	Hepatitis
42	39	78.0	11	5	ABB80558	Hepatitis
43	39	78.0	11	5	ABB80537	Hepatitis
44	39	78.0	11	5	ABB80560	Hepatitis
45	39	78.0	11	5	ABB80527	Hepatitis

ALIGNMENTS

RESULT 1
ABB80548
ID ABB80548 standard; peptide; 11 AA.

AC ABB80548;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
XX
KW Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide; virucide.
XX
OS Synthetic.

XX
FH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 9
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunk TK;
WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.
Claim 17; Page 65; 69pp; English.
The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.012;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EEWVPXGXDYS 11
 Db 1 EEWVPXGTDYS 11

RESULT 2
 ABB80547
 ID ABB80547 standard; peptide; 11 AA.
 XX AC ABB80547;
 DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW Synthetic.
 XX OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 XX WO200208251-A2.

31-JAN-2002.
 19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.
 (CORV-) CORVAS INT INC.
 Lim-Wilby M, Levy OE, Brunck TK;
 WPI; 2002-361643/39.
 Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.
 Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus
 SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.012;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EEWVPXGXDYS 11
 Db 1 EEWVPXGTDYS 11

RESULT 3
 ABB80556
 ID ABB80556 standard; peptide; 11 AA.
 XX AC ABB80556;
 DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #36.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW Synthetic.
 XX OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6
 FT Modified-site /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 XX WO200208251-A2.

31-JAN-2002.
 19-JUL-2001; 2001WO-US023169.
 21-JUL-2000; 2000US-0220101P.
 (CORV-) CORVAS INT INC.
 Lim-Wilby M, Levy OE, Brunck TK;
 WPI; 2002-361643/39.
 Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C virus
 protease.

Claim 17; Page 65; 69pp; English.
 The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus
 SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.012;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EEWVPXGXDYS 11
 Db 1 EEWVPXGTDYS 11

RESULT 4

ABB80557
ID ABB80557 standard; peptide; 11 AA.
XX
AC ABB80557;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #37.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSDYS 11
DB 1 EEVVPXGSDYS 11
RESULT 5
ABB80551
ID ABB80551 standard; peptide; 11 AA.
XX
AC ABB80551;
XX
DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US023169.
XX
PR 21-JUL-2000; 2000US-0220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSDYS 11
DB 1 EEVVPXGSDYS 11
RESULT 6
ABB80534
ID ABB80534 standard; peptide; 11 AA.
XX
AC ABB80534;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #14.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX

```

FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site 9 residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
SQ
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 BEVVPXGXDYS 11
Db 1 BEVVPXGXDYS 11

RESULT 7
ABB80561
ID ABB80561 standard; peptide; 11 AA.
XX ABB80561;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site 8 residue 7"
FT Misc-difference 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"

```

```

FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;
SQ
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 BEVVPXGXDYS 11
Db 1 BEVVPXGMDYS 11

RESULT 8
ABB80542
ID ABB80542 standard; peptide; 11 AA.
XX ABB80542;
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site 8 residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.

```

XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX PS WPI; 2002-361643/39.
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGQDYS 11
RESULT 9
ID ABB80543 standard; peptide; 11 AA.
XX ABB80543;
XX 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "D-form residue"
FT Misc-difference 8 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.

XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX CC invention are alpha-ketoamide peptide analogues. The peptides have
XX CC virucide activity, and are useful for treating and in the manufacture of
XX CC a medicament to treat disorders associated with HCV protease. A
XX CC pharmaceutical composition comprising the peptide as an active ingredient
XX CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGQDYS 11
RESULT 10
ID ABB80524 standard; peptide; 11 AA.
XX ABB80524;
XX 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX PT activity useful for treating disorders associated with hepatitis C virus
XX PT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having

XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

XX CC invention are alpha-ketoamide peptide analogues. The peptides have

XX CC virucide activity, and are useful for treating and in the manufacture of

XX CC a medicament to treat disorders associated with HCV protease. A

XX CC pharmaceutical composition comprising the peptide as an active ingredient

XX CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXNDYS 11
|||||

Db 1 EEVVPXGWDYS 11
|||||

RESULT 11

ABB80533

ID ABB80533 standard; peptide; 11 AA.

AC ABB80533;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #13.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX KW virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

XX PT activity useful for treating disorders associated with hepatitis C virus

XX PT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having

XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

XX CC invention are alpha-ketoamide peptide analogues. The peptides have

XX CC virucide activity, and are useful for treating and in the manufacture of

XX CC a medicament to treat disorders associated with HCV protease. A

XX CC pharmaceutical composition comprising the peptide as an active ingredient

XX CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXNDYS 11
|||||

Db 1 EEVVPXGWDYS 11
|||||

RESULT 12

ABB80529

ID ABB80529 standard; peptide; 11 AA.

AC ABB80529;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

XX KW virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with

FT Modified-site 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US023169.

XX PR 21-JUL-2000; 2000US-0220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-Wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

XX PT activity useful for treating disorders associated with hepatitis C virus

XX PT protease.

XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having

XX CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

XX CC invention are alpha-ketoamide peptide analogues. The peptides have

XX CC virucide activity, and are useful for treating and in the manufacture of

XX CC a medicament to treat disorders associated with HCV protease. A

XX CC pharmaceutical composition comprising the peptide as an active ingredient

XX CC is useful for treating disorders associated with hepatitis C virus

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB ||||| |||||
1 EEVVPXGMDYS 11

RESULT 13

ABB80528
ID ABB80528 standard; peptide; 11 AA.

AC ABB80528;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XW virucide.
XX Synthetic.

FS Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 8 residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.

PN 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US023169.
XX 21-JUL-2000; 2000US-0220101P.
XX (CORV-) CORVAS INT INC.
PI Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitor
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB ||||| |||||
1 EEVVPXGMDYS 11

RESULT 14

ABB80538

ID ABB80538 standard; peptide; 11 AA.

AC ABB80538;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XW virucide.
XX Synthetic.

FS Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 9 residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US023169.

XX 21-JUL-2000; 2000US-0220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-Wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB ||||| |||||
1 EEVVPXGMDYS 11

RESULT 15

ABB80562
ID ABB80562 standard; peptide; 11 AA.

AC ABB80562;
DT 08-OCT-2002 (first entry)
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.
XX

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 OS Synthetic.
 XX
 XX
 XX Key Location/Qualifiers
 FT Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8
 FT /note= "Oxymethionine"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 XX WO200208251-A2.
 PN
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US023169.
 PF
 XX
 XX 21-JUL-2000; 2000US-0220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX
 XX Lim-Wilby M, Levy OE, Bruck TK;
 PI
 XX
 XX WPI; 2002-361643/39.
 DR
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 PT
 XX
 PS Claim 17; Page 65; 69pp; English.
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus
 XX
 SQ Sequence 11 AA;
 Query Match 92.0%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXGYS 11
 Db 1 EEVVPXGXGYS 11
 Search completed: June 3, 2004, 11:48:25
 Job time : 46.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:36:47 ; Search time 11.7333 Seconds
(without alignments)
48.399 Million cell updates/sec

Title: US-09-909-164-46

Perfect score: 50

Sequence: 1 BEVVPXGXDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	37	74.0	181	4	US-09-134-000C-4848
2	33	66.0	140	3	US-08-569-147-76
3	33	66.0	140	3	US-08-569-147-82
4	32	64.0	139	4	US-09-134-001C-5124
5	32	64.0	301	4	US-09-830-217-14
6	32	64.0	3472	4	US-09-408-020-4
7	31	62.0	303	4	US-09-134-000C-4318
8	31	62.0	622	2	US-08-458-146-2
9	31	62.0	622	2	US-08-459-085-2
10	31	62.0	739	3	US-09-413-814-86
11	30.5	61.0	669	4	US-09-617-594A-2
12	30	60.0	59	4	US-08-963-851-14
13	30	60.0	120	4	US-09-621-976-4604
14	30	60.0	121	4	US-09-152-060-68
15	30	60.0	121	4	US-09-152-060-85
16	30	60.0	122	2	US-08-679-995A-1
17	30	60.0	122	3	US-09-215-096-1
18	30	60.0	217	4	US-09-543-681A-5172
19	30	60.0	231	3	US-08-926-842B-20
20	30	60.0	240	3	US-08-926-842B-21
21	30	60.0	241	3	US-08-834-776A-2
22	30	60.0	341	3	US-08-853-948B-4
23	30	60.0	348	3	US-08-853-948B-5
24	30	60.0	368	4	US-09-697-367-24
25	30	60.0	421	3	US-09-093-448-4
26	30	60.0	421	4	US-09-813-555-4
27	30	60.0	421	4	US-09-523-263B-18

28	30	60.0	425	4	US-09-252-991A-25088	Sequence 25088, A
29	30	60.0	478	2	US-09-040-799-3	Sequence 3, Appl1
30	30	60.0	478	3	US-09-093-448-1	Sequence 1, Appl1
31	30	60.0	478	3	US-09-093-448-2	Sequence 2, Appl1
32	30	60.0	478	3	US-09-093-448-3	Sequence 3, Appl1
33	30	60.0	478	4	US-09-813-555-1	Sequence 1, Appl1
34	30	60.0	478	4	US-09-813-555-2	Sequence 2, Appl1
35	30	60.0	478	4	US-09-813-555-3	Sequence 3, Appl1
36	30	60.0	478	4	US-09-523-263B-4	Sequence 4, Appl1
37	30	60.0	478	4	US-09-523-263B-16	Sequence 16, Appl1
38	30	60.0	478	4	US-09-523-263B-17	Sequence 17, Appl1
39	30	60.0	480	4	US-09-252-991A-22164	Sequence 22164, A
40	30	60.0	864	4	US-09-810-268-3	Sequence 3, Appl1
41	30	60.0	1049	4	US-09-394-272-10	Sequence 10, Appl1
42	30	60.0	1068	2	US-08-429-054A-11	Sequence 11, Appl1
43	30	60.0	1068	2	US-08-718-777-7	Sequence 7, Appl1
44	30	60.0	1068	3	US-09-051-341-7	Sequence 7, Appl1
45	30	60.0	1068	4	US-09-394-272-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-09-134-000C-4848
; Sequence 4848, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 03796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4848
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4848

Query Match 74.0%; Score 37; DB 4; Length 181;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
DB 145 BEVVPXGXDY 154

RESULT 2
US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-76

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXDY 10
Db 122 VVPTGFDY 129

RESULT 3
US-08-569-147-82
Sequence 82, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESS: No. 6180377x1s, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-82

Query Match 65.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXDY 10
Db 122 VVPTGFDY 129

RESULT 4
US-09-134-001C-5124
Sequence 5124, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5124
LENGTH: 139
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5124

Query Match 64.0%; Score 32; DB 4; Length 139;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXDYS 11
Db 32 IVPFGHDYN 40

RESULT 5
US-09-830-217-14
Sequence 14, Application US/09830217
Patent No. 6521441
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
FILE REFERENCE: PB461PCT
CURRENT APPLICATION NUMBER: US/09/830,217
CURRENT FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: PCT/US99/06199
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/080,296
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/084,674
PRIOR FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 301
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-830-217-14

Query Match 64.0%; Score 32; DB 4; Length 301;
Best Local Similarity 55.6%; Pred. No. 92;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXDYS 11
Db 195 IVPFGHDYN 203

RESULT 6
US-09-408-020-4
Sequence 4, Application US/09408020
Patent No. 6632937

```

; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 64.0%; Score 32; DB 4; Length 3472;
Best Local Similarity 45.5%; Pred. NO. 1.5e+03;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGDY 11
Db 2294 EDVIPRGISFS 2304

; RESULT 7
US-09-134-000C-4318
; Sequence 4318, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4318
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (2)...(2)
; OTHER INFORMATION: Amino acid 2 is Xaa wherein Xaa = any amino acid.
US-09-134-000C-4318

Query Match 62.0%; Score 31; DB 4; Length 303;
Best Local Similarity 40.0%; Pred. NO. 1.5e+02;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGDY 10
Db 222 EQITPTGIEY 231

; RESULT 8
US-08-459-146-2
; Sequence 2, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; HYPORULANCE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.

```

```

; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptonectria)
; ORGANISM: parasitica
; STRAIN: EP713
US-08-459-146-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. NO. 3.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXG 7
Db 31 EEVVPAG 37

; RESULT 9
US-08-459-065-2
; Sequence 2, Application US/08459065
; Patent No. 5882642
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; HYPORULANCE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,065
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptosporidia)
; STRAIN: EP713
US-08-459-065-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXG 7
DB 31 EEVVPAG 37

RESULT 10
US-09-413-814-86
; Sequence 86, Application US/09413814
; Patent No. 6225564
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 86
; TYPE: PRT
; LENGTH: 739
; ORGANISM: Sorangium cellulosum
US-09-413-814-86

Query Match 62.0%; Score 31; DB 3; Length 739;
Best Local Similarity 62.5%; Pred. No. 4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGXDYS 11
DB 663 IPLGDYS 670

RESULT 11
US-09-617-594A-2
; Sequence 2, Application US/09617594A
; Patent No. 6541458
; GENERAL INFORMATION:
; APPLICANT: Audonnet, et al.
; TITLE OF INVENTION: FELINE CALICIVIRUS GENES AND VACCINES IN PARTICULAR RECOMBINANT
; CURRENT APPLICATION NUMBER: US/09/617,594A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/193,332
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: France 00 01761
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: France 99 09421
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Feline calicivirus
US-09-617-594A-2

Query Match 61.0%; Score 30.5; DB 4; Length 669;
Best Local Similarity 63.6%; Pred. No. 4.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EEVVPXGXDYS 11
DB 428 BELTPAG-DYS 437

RESULT 12
US-08-963-851-14
; Sequence 14, Application US/08963851
; Patent No. 6300116
; GENERAL INFORMATION:
; APPLICANT: VAN DER OSTEN, CLAUS
; APPLICANT: HALKIER, TORDEN
; APPLICANT: ANDERSEN, CARSTEN
; APPLICANT: BAUDITZ, PETER
; APPLICANT: HANSEN, PETER KAMP
; TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
; FILE REFERENCE: 4946,200-US
; CURRENT APPLICATION NUMBER: US/08/963,851
; CURRENT FILING DATE: 1997-11-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-08-963-851-14

Query Match 60.0%; Score 30; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 38 EXHIPGGLYS 48

RESULT 13
US-09-621-976-4604
; Sequence 4604, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
```

; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4604
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4604

Query Match 60.0%; Score 30; DB 4; Length 120;
Best Local Similarity 50.0%; Pred. No. 82;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVXPXGXDYS 11
|:|:|:|:|:|
Db 37 EILPSSDRS 46

RESULT 14
US-09-152-060-68
; Sequence 68, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-152-060-68

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXG 7
|:|:|:|:|
Db 28 EEVVPXG 34

RESULT 15
US-09-152-060-85
; Sequence 85, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US

; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-152-060-85

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXG 7
|:|:|:|:|
Db 28 EEVVPXG 34

Search completed: June 3, 2004, 12:03:09
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 BEVVPXGXDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	94.0	11	12	US-09-909-164-31
2	47	94.0	11	12	US-09-909-164-32
3	47	94.0	11	12	US-09-909-164-35
4	47	94.0	11	12	US-09-909-164-40
5	47	94.0	11	12	US-09-909-164-41
6	46	92.0	11	12	US-09-909-164-8
7	46	92.0	11	12	US-09-909-164-12
8	46	92.0	11	12	US-09-909-164-17
9	46	92.0	11	12	US-09-909-164-18
10	46	92.0	11	12	US-09-909-164-22
11	46	92.0	11	12	US-09-909-164-26
12	46	92.0	11	12	US-09-909-164-27
13	46	92.0	11	12	US-09-909-164-45
14	46	92.0	11	12	US-09-909-164-45
15	46	92.0	11	12	US-09-909-164-46

16	41	82.0	11	12	US-09-909-164-28
17	41	82.0	11	12	US-09-909-164-29
18	41	82.0	11	12	US-09-909-164-33
19	41	82.0	11	12	US-09-909-164-36
20	41	82.0	11	12	US-09-909-164-37
21	40	80.0	11	12	US-09-909-164-5
22	40	80.0	11	12	US-09-909-164-6
23	40	80.0	11	12	US-09-909-164-9
24	40	80.0	11	12	US-09-909-164-10
25	40	80.0	11	12	US-09-909-164-14
26	40	80.0	11	12	US-09-909-164-19
27	40	80.0	11	12	US-09-909-164-20
28	40	80.0	11	12	US-09-909-164-23
29	40	80.0	11	12	US-09-909-164-24
30	40	80.0	11	12	US-09-909-164-30
31	40	80.0	11	12	US-09-909-164-34
32	40	80.0	11	12	US-09-909-164-38
33	40	80.0	11	12	US-09-909-164-39
34	40	80.0	11	12	US-09-909-164-43
35	40	80.0	11	12	US-09-909-164-47
36	40	80.0	11	12	US-09-909-164-48
37	40	80.0	11	12	US-09-909-164-49
38	40	80.0	11	12	US-09-909-164-50
39	40	80.0	11	12	US-09-909-164-51
40	40	80.0	11	12	US-09-909-164-52
41	39	78.0	11	12	US-09-909-164-7
42	39	78.0	11	12	US-09-909-164-11
43	39	78.0	11	12	US-09-909-164-15
44	39	78.0	11	12	US-09-909-164-16
45	39	78.0	11	12	US-09-909-164-21

ALIGNMENTS

RESULT 1

US-09-909-164-31
; Sequence 31, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)-(11)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)-(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)-(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-31

Query Match 94.0%; Score 47; DB 12; Length 11;

```

; Best Local Similarity 90.9%; Pred.No. 0.0049;
; Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGXSDYS 11
        |||||
Db       1 BEVVPXGTDYS 11

RESULT 2
US-09-909-164-32
; Sequence 32, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match          94.0%   Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.0049;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGXSDYS 11
        |||||
Db       1 BEVVPXGSDYS 11

RESULT 4
US-09-909-164-40
; Sequence 40, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match          94.0%   Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred.No. 0.0049;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGXSDYS 11
        |||||
Db       1 BEVVPXGTDYS 11

RESULT 3
US-09-909-164-35
; Sequence 35, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21

```

```
; OTHER INFORMATION: D-amino acid
US-09-909-164-40
Query Match          94.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0049;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGSDYS 11

RESULT 5
US-09-909-164-41
; Sequence 41, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-8
;
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGMDYS 11

RESULT 7
US-09-909-164-12
; Sequence 12, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
```

```
; OTHER INFORMATION: D-amino acid
US-09-909-164-40
Query Match          94.0%; Score 47; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0049;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGSDYS 11

RESULT 5
US-09-909-164-41
; Sequence 41, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-8
;
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGMDYS 11

RESULT 7
US-09-909-164-12
; Sequence 12, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
```



```
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-12
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
    |||||
Db 1 BEVVPXGMDYS 11
    |||||

RESULT 8
US-09-909-164-13
; Sequence 13, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-13
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
    |||||
Db 1 BEVVPXGMDYS 11
    |||||

RESULT 9
US-09-909-164-17
; Sequence 17, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
US-09-909-164-13
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
    |||||
Db 1 BEVVPXGMDYS 11
    |||||

RESULT 10
US-09-909-164-18
; Sequence 18, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
```

```
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
US-09-909-164-18
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXDYS 11
Db 1 EVVVPXGQDYS 11

RESULT 11
US-09-909-164-22
; Sequence 22, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-26
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXDYS 11
Db 1 EVVVPXGQDYS 11

RESULT 13
US-09-909-164-27
; Sequence 27, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
US-09-909-164-22
Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVVPXGXDYS 11
Db 1 EVVVPXGQDYS 11

RESULT 12
US-09-909-164-26
; Sequence 26, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
```

```

; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; US-09-909-164-27

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
   |||||
Db 1 EEVVPXGQDYS 11

RESULT 14
US-09-909-164-45
; Sequence 45, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; US-09-909-164-46

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
   |||||
Db 1 EEVVPXGXDYS 11

US-09-909-164-45
; Sequence 45, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; US-09-909-164-45

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
   |||||
Db 1 EEVVPXGXDYS 11

```

```

RESULT 15
US-09-909-164-46
; Sequence 46, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: Met (O)
; US-09-909-164-46

Query Match          92.0%; Score 46; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
   |||||
Db 1 EEVVPXGXDYS 11

Search completed: June 3, 2004, 12:57:17
Job time : 33.7333 secs

```

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 BEVVPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	72.0	363	2 D69551	conserved hypothetical
2	35	70.0	290	2 AG3104	6-O-methylguanine
3	35	70.0	290	2 D98182	O6-methylguanine-D
4	35	70.0	587	2 F81138	succinate dehydrog
5	35	70.0	906	2 T48898	disease resistance
6	35	70.0	908	2 T48899	disease resistance
7	34	68.0	102	2 A42452	V1 protein - tobac
8	34	68.0	156	2 S54619	hypothetical prote
9	34	68.0	247	2 A96001	conserved hypothet
10	34	68.0	257	2 A96546	unknown protein [1
11	34	68.0	394	2 F82491	ferrisiderophore r
12	34	68.0	433	2 H87660	peptidoglycan-bind
13	34	68.0	2747	2 B49132	fat facies (fat) s
14	33	66.0	124	1 VKJTSI	trans-regulatory s
15	33	66.0	165	2 AG1272	thiol peroxidases
16	33	66.0	165	2 AH1635	thiol peroxidases
17	33	66.0	196	2 AD0454	conserved hypothet
18	33	66.0	225	2 S57810	hypothetical prote
19	33	66.0	327	2 S40753	hypothetical prote
20	33	66.0	421	1 DERTCM	acyl-CoA dehydroge
21	33	66.0	440	2 H72784	probable alkaline
22	33	66.0	1028	2 AF3286	ATP-dependent DNA
23	33	66.0	1088	2 D82246	probable chitinase
24	33	66.0	1150	2 T20173	hypothetical prote
25	32	64.0	99	2 S00210	plastocyanin b - L
26	32	64.0	155	2 S38255	plastocyanin precu
27	32	64.0	168	2 S58208	plastocyanin b pre
28	32	64.0	196	2 AI0931	conserved hypothet
29	32	64.0	301	2 F89957	hypothetical prote

30 32 64.0 307 2 F84330
31 32 64.0 314 2 AH1912
32 32 64.0 357 1 G59290
33 32 64.0 366 2 G59350
34 32 64.0 425 2 T24111
35 32 64.0 427 2 F64064
36 32 64.0 565 2 E86665
37 32 64.0 632 2 H84350
38 32 64.0 672 2 G88651
39 32 64.0 1474 2 F69009
40 32 64.0 2472 2 T31308
41 32 64.0 6658 2 TL3931
42 31 62.0 117 2 A69487
43 31 62.0 202 2 H97247
44 31 62.0 233 2 E72330
45 31 62.0 296 2 F72745

hypothetical prote
hypothetical prote
probable hexosyltr
L-lactate dehydrog
hypothetical prote
ABC transporter AT
oligopeptidase [im
protein B0212.3 [i
probable membrane
hypothetical 367K
projectin - fruit
response regulator
S-adenosylmethioni
conserved hypothet
hypothetical prote

conserved hypothetical protein AF2411 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: D69551
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69551
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-363 <KLE>
A:Cross-references: GB:AE001109; GB:AE000782; NID:g26689432; PIDN:AA891255.1; PID:g265061

Query Match 72.0%; Score 36; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
Db 120 ENIVPYGIDFS 130

RESULT 2
AG3104
6-O-methylguanine-DNA methyltransferase [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG3104
R;Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Foo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KUT>
A:Cross-references: GB:AE008689; PIDN:AAU45253.1; PID:gl7742937; GSPDB:GN00187
C:Genetics:
A:Gene: ada
A:Map position: linear chromosome

ALIGNMENTS

RESULT 1

```

Query Match      70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
   |::|||
Db 9 EDITPIGSDY 18

RESULT 3
O6-methylguanine-DNA methyltransferase PA2118 [imported] - Agrobacterium tumefaciens (strain D98182)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: D98182
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens D98182
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D98182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88982.1; PID:G15158766; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L 818
A:Map position: linear chromosome

Query Match      70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
   |::|||
Db 9 EDITPIGSDY 18

RESULT 4
F81138
succinate dehydrogenase, flavoprotein chain NMB0950 [imported] - Neisseria meningitidis C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81138
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickley, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: F81138
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-587 <TET>
A:Cross-references: GB:AE002446; GB:AE002098; NID:G7226185; PIDN:AAF41356.1; PID:G722618
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0950
C:Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; F81138

Query Match      70.0%; Score 35; DB 2; Length 587;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
   |::|||
Db 366 EVVVPQGEDY 375

RESULT 5
T48899
disease resistance protein RPP8 [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48899
R:McDowell, J.M.; Dhandaaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Danjani, P.; Title: Intragenic recombination and diversifying selection contribute to the evolution of the RPP8 gene in Arabidopsis thaliana.
A:Reference number: Z24999; MUID:99030193; PMID:9811794
A:Accession: T48899
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-906 <MCD>
A:Cross-references: EMBL:AF089710; NID:G3928861; PIDN:AAK83165.1; PID:G3928862
A:Experimental source: Landsberg erecta
C:Genetics:
A:Gene: RPP8
A:Introns: 293/1; 342/1
A:Description: promotes resistance to Peronospora parasitica
A:Function: promotes resistance to Peronospora parasitica

Query Match      70.0%; Score 35; DB 2; Length 906;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
   |::|||
Db 881 EKLVPQGEDY 890

RESULT 6
T48899
disease resistance protein rpp8 [similarity] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48899
R:McDowell, J.M.; Dhandaaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Danjani, P.; Title: Intragenic recombination and diversifying selection contribute to the evolution of the RPP8 gene in Arabidopsis thaliana.
A:Reference number: Z24999; MUID:99030193; PMID:9811794
A:Accession: T48899
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-908 <MCD>
A:Cross-references: EMBL:AF089711; NID:G3901293; PIDN:AAK78631.1; PID:G3901294
A:Experimental source: Columbia
C:Genetics:
A:Gene: rpp8
A:Introns: 293/1; 342/1
A:Function: promotes resistance to Peronospora parasitica
A:Description: susceptible allele of a gene that promotes resistance to Peronospora parasitica

Query Match      70.0%; Score 35; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
   |::|||
Db 883 EKLVPQGEDY 892

RESULT 7
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E. Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellows virus.
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>

```

A:Cross-References: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 68.0%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 6.1;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

QY 2 EVVVPXGXDYS 11
:|:|:|:|:|:
DB 7 QVVPSSGINS 16

RESULT 8

S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S54619; S66879
R:De Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54617
A:Accession: S54619
A:Molecule type: DNA
A:Residues: 1-156 <DEH>
A:Cross-References: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123
R:De Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66877
A:Accession: S66879
A:Molecule type: DNA
A:Residues: 1-156 <DEW>
A:Cross-References: EMBL:Z74920; NID:G1420109; PIDN:CAA99201.1; PID:G1420111; MIPS:YOR01
A:Experimental source: strain S288C
C:Genetics:
A:Cross-References: SGD:S0005539
A:Map position: 15R
C:Superfamily: hypothetical protein YOR013w

Query Match 68.0%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 9.6;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 2 EVVVPXGXDY 10
:|:|:|:|:|:
DB 50 EVWPLGMDY 58

RESULT 9

A96001
conserved hypothetical protein, homolog to osmotically inducible sensory protein SMC22-1
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: A96001
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-Kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A96001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-247 <KUR>
A:Cross-References: GB:AL591985; PIDN:CA49673.1; PID:G15141160; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
nebaut, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:

A:Gene: SMD21444
A:Genome: plasmid

Query Match 68.0%; Score 34; DB 2; Length 247;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 1 BEVVPXGXDY 10
:|:|:|:|:|:
DB 48 EDVEPRGADY 57

RESULT 10

A96546
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96546
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <SNO>
A:Cross-References: GB:AE005173; NID:G11094688; PIDN:AAG29624.1; GSPDB:GN00141
C:Genetics:
A:Gene: F8A12.12
A:Map position: 1

Query Match 68.0%; Score 34; DB 2; Length 257;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

QY 1 BEVVPXGXDS 11
:|:|:|:|:|:
DB 217 BELKAGADYS 227

RESULT 11

F82491
ferrisiderophore reductase VCA0183 [imported] - Vibrio cholerae (strain N16961 serogrou
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82491
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82491
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <HEI>
A:Cross-References: GB:AE004358; GB:AE003853; NID:G9657566; PIDN:AAF96096.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0183
A:Map position: 2
C:Superfamily: flavohemoglobin; cytochrome-b5 reductase homology; globin homology

Query Match 68.0%; Score 34; DB 2; Length 394;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs prote
 C;Species: simian immunodeficiency virus SIVcpz
 A;Note: host Pan troglodytes (chimpanzee)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: S09988
 R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A;Reference number: S09988; MUID:90259077; PMID:2188136
 A;Accession: S09988
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-124 <HUE>
 A;Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
 C;Genetics:
 A;Gene: rev; trs; art
 A;Introns: 27/1
 C;Superfamily: AIDS trans-regulatory splicing protein
 C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 66.0%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXSDY 11
 DB 107 ETVFAGGNY 116

RESULT 15
 AG1272
 thiol peroxidases homolog lmo1583 [imported] - Listeria monocytogenes (strain EGD-e)
 C;Species: Listeria monocytogenes
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C;Accession: AG1272
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
 D.; Jones, L.M.; Karst, J.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Maduano, E.; Maitournam, A.; M
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
 A;Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AG1272
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-165 <GLA>
 A;Cross-references: GB:NC_003210; PIDN:CAC99661.1; PID:G16411012; GSPDB:GN00177
 A;Experimental source: strain EGD-e
 C;Genetics:
 A;Gene: lmo1583
 C;Superfamily: thioredoxin peroxidase

Query Match 66.0%; Score 33; DB 2; Length 165;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSDY 10
 DB 144 EVVPEGSDH 152

Search completed: June 3, 2004, 12:00:03
 Job time : 9 secs

N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs prote
 C;Species: simian immunodeficiency virus SIVcpz
 A;Note: host Pan troglodytes (chimpanzee)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: S09988
 R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A;Reference number: S09988; MUID:90259077; PMID:2188136
 A;Accession: S09988
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-124 <HUE>
 A;Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
 C;Genetics:
 A;Gene: rev; trs; art
 A;Introns: 27/1
 C;Superfamily: AIDS trans-regulatory splicing protein
 C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 68.0%; Score 34; DB 2; Length 433;
 Best Local Similarity 54.5%; Pred. No. 30;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXSDY 11
 DB 266 EVILPFGDYS 276

RESULT 13
 B49132
 fat facets (faf) splice form 1 - fruit fly (Drosophila melanogaster)
 C;Species: Drosophila melanogaster
 C;Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 01-Dec-2000
 C;Accession: B49132; A49132
 R;Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.
 Development 116, 985-1000, 1992
 A;Title: The fat facets gene is required for Drosophila eye and embryo development.
 A;Reference number: A49132; MUID:93202020; PMID:1295547
 A;Contents: isogenic st
 A;Accession: B49132
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1-2747 <FIS>
 A;Cross-references: GB:L04959; NID:G157411; PIDN:AAF01345.1; PID:G6013474
 A;Note: sequence inconsistent with the nucleotide translation
 A;Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:127839)
 A;Accession: A49132
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1-2704, 'VT', '2707', 'ANNV', <PI2>
 A;Cross-references: GB:L04958; NID:G157410; PIDN:AAF01346.1; PID:G6013475
 A;Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:129029)
 C;Keywords: alternative splicing

Query Match 68.0%; Score 34; DB 2; Length 2747;
 Best Local Similarity 54.5%; Pred. No. 2.2e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXSDY 11
 DB 1394 EVIPDGGQDS 1404

RESULT 14
 VKLJSI
 trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 Seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 BEVVPXGXDSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	70.0	908	1 R8L4 ARATH	Q9fj18 arabidopsis
2	35	70.0	908	1 RFP8 ARATH	Q8w4j9 arabidopsis
3	35	70.0	910	1 RPH8 ARATH	P59584 arabidopsis
4	35	70.0	1058	1 CARB FUSN	Q8r986 fusobacteri
5	34	68.0	102	1 Y1LK TYDVA	P31619 tobacco yel
6	34	68.0	394	1 HMPA VIBCH	Q9kmv3 vibrio chol
7	34	68.0	2778	1 FAP BROME	P55924 drosophila
8	33	66.0	124	1 REV SVICZ	P27280 chimpanzee
9	33	66.0	165	1 TPX LISIN	Q92bc5 listeria in
10	33	66.0	165	1 TPX LISMO	Q9y6u8 listeria mo
11	33	66.0	327	1 YK14 CAEEL	P34338 caenorhabdi
12	33	66.0	421	1 ACDM RAT	P08503 rattus norv
13	33	66.0	563	1 SVR STRP3	Q8k5j2 streptococc
14	33	66.0	563	1 SVR STRP8	Q8n222 streptococc
15	33	66.0	563	1 SVR STRPY	Q9nx15 streptococc
16	32	64.0	154	1 PLAS OKISA	P20423 oryza sativ
17	32	64.0	155	1 PLAS HORVU	P08248 hordeum vul
18	32	64.0	168	1 PLAT POPNI	P11970 populus nig
19	32	64.0	196	1 YJAG SALTV	Q91912 salmonele
20	32	64.0	302	1 DDL GIOVI	O7nnw1 gloeobacter
21	32	64.0	336	1 SUGT MOUSE	Q8cx34 mus musculu
22	32	64.0	427	1 TOUB HAEIN	P44677 haemophilus
23	32	64.0	1499	1 CHAC HUMAN	Q80312 homo sapien
24	32	64.0	3174	1 CHAC HUMAN	Q96rl7 homo sapien
25	31	62.0	233	1 HIS9 THEWA	Q9wzr1 thermotoga
26	31	62.0	319	1 YHAI-CRYPA	P10941 cryphonectr
27	31	62.0	421	1 ACDM HUMAN	P11310 homo sapien
28	31	62.0	421	1 ACDM PIG	P41367 sus scrofa
29	31	62.0	562	1 TRDM PANAY	Q47861 pantoea agg
30	30	60.0	97	1 PLAS DAUCA	P20422 gausus caro
31	30	60.0	121	1 TKNK HUMAN	Q9uhio homo sapien
32	30	60.0	165	1 YJ49 ARCFU	Q28330 archaeoglob
33	30	60.0	175	1 HES3 RAT	Q04667 rattus norv

ALIGNMENTS

RESULT 1

ID	R8L4 ARATH	STANDARD;	PRT;	908 AA.
AC	Q9fj18			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable disease resistance RPP8-like protein 4.			
GN	RPP8L4 OR AT5G48620 OR K15N18.9.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=99087489; PubMed=9872454;			
RA	Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. VII.			
RT	Sequence features of the regions of 1,013,767 bp covered by sixteen			
RT	physically assigned P1 and TAC clones."			
EL	DNA Res. 5:297-308(1998).			
EL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=cv. Columbia;			
RC	STRAIN=cv. Columbia;			
RA	Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,			
RA	Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,			
RA	Havashizaki Y., Shinozaki K.;			
RT	"Arabidopsis thaliana full-length cDNA."			
RT	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
RL	-!- FUNCTION: Potential disease resistance protein.			
CC	-!- DOMAIN: The LRR repeats probably act as specificity determinant of			
CC	pathogen recognition (by similarity).			
CC	-!- SIMILARITY: Belongs to the disease resistance NB-LRR family.			
CC	RPP8/HRT subfamily.			
CC	-!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.			
CC	-!- SIMILARITY: Contains 1 NB-ARC domain.			
CC	-!- DATABASE: NAME=NIB-LRRS;			
CC	NOTE=Functional and comparative genomics of disease resistance gene			
CC	homologs;			
CC	WWW="http://niblrrs.ucdavis.edu".			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AB015468; BAB10695.1; -			
DR	EMBL; AK117163; BAC41841.1; -			
DR	InterPro; IPR000767; Disease_resist.			
DR	InterPro; IPR001611; LRR.			

P14194 bacillus su
P08203 escherichia
P06190 salmonella
Q92ie3 helicobacte
P56006 helicobacte
P06308 lymnaea sta
P38174 saccharomyc
O52250 halomonas e
Q05865 bacillus su
O15231 homo sapien
P53998 kluyveromyc

DR InterPro: IPR002182; NB-ARC.
DR Pfam: PF00560; LRR: 2.
DR Pfam: PF00931; NB-ARC: 1.
DR PRINTS: PR00364; DISEASERSIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT DOMAIN 10 45 NB-CINE-ZIPPER.
FT REPEAT 146 459
FT REPEAT 575 599 LRR 1.
FT REPEAT 600 623 LRR 2.
FT REPEAT 842 867 LRR 3.
FT NP BIND 192 199 ATP (POTENTIAL).
SQ SEQUENCE 908 AA; 104448 MW; 311991B17239693 CRC64;
Query Match 70.0%; Score 35; DB 1; Length 908;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EEVVPXGXDY 10
Db 883 EKLVPFGEDY 892
RESULT 2
RP8 ARATH
ID RP8 ARATH STANDARD; PRT; 908 AA.
AC Q8W4J9; Q8W4J9; Q8W4J9; Q8W4J9; Q8W4J9; Q8W4J9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Disease resistance protein RPP8 (Resistance to *Peronospora parasitica*
DE protein 8).
GN RPP8 OR HRT OR AT5G43470 OR MWE20.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND
RP RPP8-3, AND VARIANTS.
RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
RX MEDLINE=99030193; PubMed=9811794;
RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intragenic recombination and diversifying selection contribute to the
RT evolution of downy mildew resistance at the RPP8 locus of
RT Arabidopsis.";
RL Plant Cell 10:1861-1874(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
RC STRAIN=cv. Di-17;
RX MEDLINE=20271766; PubMed=10810142;
RA Cooley M.B., Pathirana S., Wu H.-J., Kachroo P., Klessig D.F.;
RT "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer
RT resistance to both viral and comycete pathogens.";
RL Plant Cell 12:663-676(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.F.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L.D., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Heuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RL Science 302:842-846(2003).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP INTERACTION WITH TIP.
RX MEDLINE=20496823; PubMed=11041886;
RA Ren T., Qu F., Morris T.J.;
RT "HRT gene function requires interaction between a NAC protein and
RT viral capsid protein to confer resistance to turnip crinkle virus.";
RL Plant Cell 12:1917-1926(2000).
CC -!- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. The interaction
CC with TIP (TCV-interacting protein) may be essential for the
CC recognition of the avirulence proteins, and the triggering of the
CC defense response.
CC -!- SUBUNIT: Interacts with the NAC protein TIP.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8W4J9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8W4J9-2; Sequence=VSP_007171, VSP_007172;
CC Note=Has been shown to exist only in cv. Columbia so far;
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition.
CC -!- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
CC cv. Columbia are probably due to an unequal crossing-over between
CC the highly related RPP8 and RPP8A genes present in cv. Landsberg
CC erecta. Such variations probably modify the specificity of
CC pathogen recognition.
CC -!- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
CC specifically recognizes the EmcoS avirulence protein from
CC *Peronospora parasitica*, while it is not the case in cv. Di-17,
CC where it confers resistance to Turnip Crinkle Virus upon
CC recognition of the viral capsid protein.
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- DATABASE: NAME=NB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrns.ucdavis.edu".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF089710; AAC83165.1; -.

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
 DB 885 EKLVPGGEDY 894

RESULT 4
 CARB_FUSNN STANDARD; PRT; 1058 AA.
 ID CARB_FUSNN STANDARD; PRT; 1058 AA.
 AC Q8SG86;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 GN CARB OR FN0422.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 CC Bacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 25586;
 RC MEDLINE=21886394; PubMed=11899109;
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Barton A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R., Forstein N., Kyrides N., Overbeek R.;
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";
 RT J. Bacteriol. 184:2005-2018 (2002).
 RL -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- PATHWAY: Pyrimidine biosynthesis;
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
 CC -!- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE010554; AAL94625.1; ALT_INIT.
 DR HAMAP; MF_01210; 1.
 DR InterPro; IPR006275; CARB_L_glu.
 DR InterPro; IPR005483; Cbase_L_D2.
 DR InterPro; IPR005479; Cbase_L_D2.
 DR InterPro; IPR005480; Cbase_L_D2.
 DR InterPro; IPR005481; Cbase_L_N.
 DR InterPro; IPR004362; MGS like.
 DR Pfam; PF00289; Cbase_L_Chain; 2.
 DR Pfam; PF02786; Cbase_L_D2; 2.
 DR Pfam; PF02787; Cbase_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PRO0098; CPSASE.
 DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
 DR PROSITE; PS00866; CPSASE 1; 2.
 DR PROSITE; PS00867; CPSASE 2; 2.
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.

FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 302 352 ATP (POTENTIAL).
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF7C1E39F CRC64;

Query Match 70.0%; Score 35; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred.No. 22;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSDYS 11
 DB 190 EIVPENGLYNS 199

RESULT 5
 Y11K TYDVA STANDARD; PRT; 102 AA.
 ID Y11K TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Y1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 OX NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92188538; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RA "The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants.";
 RT Virology 187:633-642(1992).
 RL This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M81103; AAA47947.1; --
 DR PIR; A42452; A42452.
 DR InterPro; IPR002621; Geminl_mov.
 DR Pfam; PF01708; Geminl_mov; 1.
 DR Hypothetical protein.
 KW SEQUENCE 102 AA; 11178 MW; A40BCF1E0AF55B67 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred.No. 3.1;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXSDYS 11
 DB 7 QVVPFGINYS 16

RESULT 6
 HMPA VIBCH STANDARD; PRT; 394 AA.
 ID HMPA VIBCH STANDARD; PRT; 394 AA.
 AC Q9NMW3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin).
 GN RMP OR VCA0183.
 GN Vibrio cholerae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 DN NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406633; PubMed=10953301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson L.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT cholerae";
 RT Nature 406:477-483(2000).
 RL -!- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
 CC OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
 CC FAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION.
 CC -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO-DOMAIN
 CC FLAVOHEMOPROTEINS SUBFAMILY.
 CC -!- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
 CC OXIDOREDUCTASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE004358; AAF96096.1; --
 DR PIR; F82491; F82491.
 DR HSSP; P39662; 1COX.
 DR TIGR; VCA0183; --
 DR InterPro; IPR008333; FAD binding_6.
 DR InterPro; IPR001709; FPN_cyt_redctse.
 DR InterPro; IPR000971; Globin.
 DR InterPro; IPR001433; Oxred_FAD/NAD(P).
 DR InterPro; IPR001221; Phe_hydroxylase.
 DR Pfam; PF00970; FAD binding_6; 1.
 DR Pfam; PF00042; Globin; 1.
 DR Pfam; PF00175; NAD binding_1; 1.
 DR PRINTS; PR00371; FPNCR.
 DR PRINTS; PR00410; PHEHYDRLASE.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport;
 KW Oxygen transport; Transports; Complete proteome.
 FT DOMAIN 1 136
 FT METAL 53 53
 FT (BY SIMILARITY)
 FT IRON (HEME DISTAL LIGAND)
 FT METAL 85 85
 FT IRON (HEME PROXIMAL LIGAND)
 FT (BY SIMILARITY)
 FT NP_BIND 268 273
 FT NADP (RIBOSE PART) (BY SIMILARITY).
 SQ SEQUENCE 394 AA; 44191 MW; DDA3490FAE28823A CRC64;
 Query Match 68.0%; Score 34; DB 1; Length 394;
 Best Local Similarity 66.7%; Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPGXDY 10
 DB 194 EVTPGSDY 202
 RESULT 7
 FAF DROME STANDARD; PRT; 2778 AA.
 ID FAF_DROME

AC P55824; Q9V9T6; Q9Y0Z7;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
 DE FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
 DE FAF OR BCDNA:ID22582 OR CG1945.
 GN Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Eye Imaginal disk;
 RX MEDLINE=93202020; PubMed=1295747;
 RA Fischer-Vize J.A., Rubin G.M., Lehmann R.;
 RA "The fat facets gene is required for Drosophila eye and embryo
 RT development";
 RT Development 116:985-1000(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Heit G., Nelson C.R., Miklos G.B.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beson K.Y., Benos P.V., Berwan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Diez S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.A.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Pan S.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=2426069; PubMed=12957572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;

RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4].
 RC SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).
 RP STRAIN=Berkley;
 RX MEDLINE=20196012; PubMed=10731138;
 RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
 RA Stapleton M., Harvey D.A.;
 RT "A Drosophila complementary DNA resource.";
 RL Science 287:2222-2224(2000).
 CC -1- FUNCTION: Required for eye and embryo development, and plays a
 CC role in compound eye assembly and oogenesis respectively, in the
 CC larval eye disks, cells outside the assembling facets require this
 CC protein for short-range cell interactions that prevent the mystery
 CC cells from becoming photoreceptors. It is also required for
 CC nuclear migration and cellularization in early embryogenesis and
 CC could play a role in pole cell determination, development or
 CC function.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=P55824-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P55824-2; Sequence=VSP_005270;
 CC Name=3;
 CC IsoId=P55824-3; Sequence=VSP_005269;
 CC -1- TISSUE SPECIFICITY: Eye disks and ovaries.
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -1- SIMILARITY: Belongs to peptidase family C19.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04959; AAF01345.1; -;
 CC EMBL; L04958; AAF01346.1; -;
 CC EMBL; L04960; AAF01347.1; -;
 CC EMBL; L04960; AAF01348.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AF145677; AAN14291.1; -;
 CC MEROPS; C19.007; AAD38652.1; -;
 CC Flybase; FBgn0005632; faf.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0007349; P:cellularization; IMP.
 CC GO; GO:0007955; P:embryonic morphogenesis; IMP.
 CC GO; GO:0007456; P:eye morphogenesis (sensu Drosophila); IMP.
 CC GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. .); IMP.
 CC GO; GO:0007097; P:nuclear migration; IMP.
 CC GO; GO:0016579; P:protein deubiquitination; IDA.
 CC GO; GO:0016511; P:ubiquitin-dependent protein catabolism; IGI.
 CC InterPro; IPR001394; Peptidase_C19.
 CC Pfam; PF00443; UCH; 1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS02035; UCH_2_3; 1.
 CC Ubl conjugation pathway; Hydrolase; Thiol protease;
 CC developmental protein; Vission; Alternative splicing.
 CC ACT SITE 1677 1677 BY SIMILARITY.
 CC ACT SITE 1978 1978 BY SIMILARITY.
 CC ACT SITE 1986 1986 BY SIMILARITY.
 CC VARSPLIC 2705 2778 KCHRVIKKLVSKDEEDATTATTATTEVTTSPATATATA
 CC ATLEPAGMSELTTWVKNLIIQENPQAKSLQ -> VTRA
 CC NNV (in isoform 3).
 FT

FT VARSPLIC 2742 2778 /FTid=VSP_005269.
 FT IATALEPAGMSELTTWVKNLIIQENPQAKSLQ ->
 FT SORQOL (in isoform 2).
 FT /FTid=VSP_005270.
 FT E -> D (IN REF. 1).
 FT CONFLICT 234 234 T -> S (IN REF. 1; AAF01345).
 FT CONFLICT 2725 2725 T -> S (IN REF. 1; AAF01345).
 SQ SEQUENCE 2778 AA; 311139 MW; FFB90438BA53A02B CRC64;
 Query Match 68.0%; Score 34; DB 1; Length 2778;
 Best Local Similarity 54.5%; Pred. No. 97;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEWVPGXGDYS 11
 DB 1394 EVIVPDGQDFS 1404
 RESULT 8
 REV_SIVCZ STANDARD; PRT; 124 AA.
 ID REV_SIVCZ STANDARD; PRT; 124 AA.
 AC P17280;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE REV protein (Anti-repression transactivator protein) (ART/TRS).
 GN REV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90259077; PubMed=2188136;
 RA Ruet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
 RL Nature 345:356-359(1990).
 CC -1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
 CC -1- PTM: Phosphoprotein whose state of phosphorylation is mediated by
 CC a specific serine kinase activity present in the nucleus.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52154; CAA36405.1; -;
 CC PIR; S09988; VKLJ51.
 CC HIV; X52154; REVSCP.
 CC InterPro; IPR000425; REV_protein.
 CC Pfam; PF00424; REV; 1.
 CC Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDF65A7B2 CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 6.2;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EEWVPGXGDYS 11
 DB 107 ETVPAGGNYS 116
 RESULT 9
 TPX_LISIN STANDARD; PRT; 165 AA.
 ID TPX_LISIN STANDARD; PRT; 165 AA.
 AC Q92BC5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable thiol peroxidase (EC 1.11.1.-).
 EN TPX OR L1N1625.
 OS Listeria innocua.
 DC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 DX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP.11262 / Serovar 5a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RA "Comparative genomics of *Listeria* species.",
 RT Science 294:849-852(2001).
 RL -!- FUNCTION: Has antioxidant activity. Could remove peroxides or
 CC H(2)O(2) (By similarity).
 CC -!- SIMILARITY: Belongs to the ahpc/tsa family. Tpx subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL596169; CAC96856.1; --
 CC PIR: AH1635; AH1635.
 CC Listlist; L1N01625; --
 CC HAMAP: MF_00269; -- 1.
 CC InterPro: IPR000866; Ahpc-TSA.
 CC Pfam: PF00578; Ahpc-TSA; 1.
 CC PROSITE: PS01265; TPX; 1.
 CC Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
 KW SEQUENCE 165 AA; 18162 MW; 77705B7CD8BC8F4D CRC64;
 SQ
 Query Match 66.0%; Score 33; DB 1; Length 165;
 Best Local Similarity 66.7%; Pred. No. 8.4;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 2Y 2 EVVPGXNDY 10
 DB 144 EVVPEGS DH 152
 RESULT 10
 TPX LISMO STANDARD; PRT; 165 AA.
 AC Q8Y608;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thiol peroxidase (EC 1.11.1.-).
 GN TPX OR LMO1583.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RA "Comparative genomics of *Listeria* species.",
 RT Science 294:849-852(2001).
 RL -!- FUNCTION: Has antioxidant activity. Could remove peroxides or
 CC H(2)O(2) (By similarity).
 CC -!- SIMILARITY: Belongs to the ahpc/tsa family. Tpx subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL591979; CAC99661.1; --
 CC PIR: AG1272; AG1272.
 CC Listlist; LMO01583; --
 CC HAMAP: MF_00269; -- 1.
 CC InterPro: IPR000866; Ahpc-TSA.
 CC Pfam: PF00578; Ahpc-TSA; 1.
 CC PROSITE: PS01265; TPX; 1.
 CC Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
 KW SEQUENCE 165 AA; 18133 MW; 77705B7CC46D424D CRC64;
 SQ
 Query Match 66.0%; Score 33; DB 1; Length 165;
 Best Local Similarity 66.7%; Pred. No. 8.4;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXNDY 10
 DB 144 EVVPEGS DH 152
 RESULT 11
 YK14 CAEEL STANDARD; PRT; 327 AA.
 ID YK14 CAEEL
 AC P34338;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C15H7.4 in chromosome III.
 GN C15H7.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Anscomb R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA LaReille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT elegans".
 RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 222173; CA80126.1; -;
CC PIR; S40753; S40753;
CC WormPep; C15H7.4; CE00082.
CC Hypothetical protein.
CC SEQUENCE 327 AA; 35566 MW; 716BC2BDD2E9607E CRC64;
CC
CC Query Match 66.0%; Score 33; DB 1; Length 327;
CC Best Local Similarity 63.6%; Pred. No. 17;
CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC Qy 1 BEVVPXGXDXS 11
CC :|||||
CC 175 KEVVPNGGDKS 185
CC
CC RESULT 12
CC ACDM_RAT STANDARD; PRT; 421 AA.
CC ID ACDM_RAT
CC AC P08503;
CC DT 01-AUG-1988 (Rel. 08, Created)
CC DT 01-AUG-1988 (Rel. 08, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Acyl-CoA dehydrogenase, medium-chain specific, mitochondrial precursor
CC (EC 1.3.99.3) (MCDAD).
CC GN ACDM.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
CC OX NCBI_TaxID=10116;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC RX TISSUE=Liver;
CC RC MEDLINE=87280028; PubMed=3611054;
CC RA Matsubara Y., Kraus J.P., Ozasa H., Glassberg R., Finocchiaro G.,
CC Ikeda Y., Mole J., Rosenberg L.E., Tanaka K.;
CC "Molecular cloning and nucleotide sequence of cDNA encoding the
CC RT entire precursor of rat liver medium chain acyl coenzyme A
CC RT dehydrogenase."
CC RL J. Biol. Chem. 262:10104-10108(1987).
CC CC -!- FUNCTION: This enzyme is specific for acyl chain lengths of 4 to
CC 16.
CC CC -!- CATALYTIC ACTIVITY: Acyl-CoA + ETF = 2,3-dehydroacyl-CoA + reduced
CC ETF.
CC CC -!- COFACTOR: FAD.
CC CC -!- PATHWAY: Mitochondrial fatty acid beta-oxidation system; first
CC step.
CC CC -!- SUBUNIT: Homotetramer.
CC CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC CC -!- MISCELLANEOUS: A number of straight-chain acyl-CoA dehydrogenases
CC of different substrate specificities are present in mammalian
CC tissues.
CC CC -!- SIMILARITY: Belongs to the acyl-CoA dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J02791; AAA40670.1; -;
CC PIR; A28436; DERTCM.
CC HSP; P11310; LEGD.
CC

DR InterPro; IPR006089; Acyl-CoA dh.
DR InterPro; IPR006090; Acyl-CoA dh. C.
DR InterPro; IPR006091; Acyl-CoA dh. M.
DR InterPro; IPR006092; Acyl-CoA dh. N.
DR Pfam; PF00441; Acyl-CoA dh. 1.
DR Pfam; PF02770; Acyl-CoA dh. M.; 1.
DR Pfam; PF02771; Acyl-CoA dh. N.; 1.
DR PROSITE; PS00072; ACYL_COA_DH_1; 1.
DR PROSITE; PS00073; ACYL_COA_DH_2; 1.
DR Oxidoreductase; Flavoprotein; FAD; Fatty acid metabolism;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 25
FT CHAIN 26 421
FT ACYL-CoA DEHYDROGENASE, MEDIUM-CHAIN
FT SPECIFIC.
FT ACT_SITE 193 193
FT FORMS A HYDROGEN-BOND WITH THE FLAVIN
FT N(5) OF THE FAD COFACTOR (BY SIMILARITY).
FT ACT_SITE 401 401
FT BASE (BY SIMILARITY).
SQ SEQUENCE 421 AA; 46555 MW; 2CF076FC919BDE8 CRC64;
CC
CC Query Match 66.0%; Score 33; DB 1; Length 421;
CC Best Local Similarity 50.0%; Pred. No. 22;
CC Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
CC
CC Qy 1 BEVVPXGXDX 10
CC :||:|
CC 58 EEIIPVADY 67
CC
CC Db 58 EEIIPVADY 67
CC
CC RESULT 13
CC SYR_STRP3 STANDARD; PRT; 563 AA.
CC ID SYR_STRP3
CC AC Q8K5J2;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
CC GN ARGS OR SPYM3_1809 OR SPS1807.
CC OS Streptococcus pyogenes (serotype M3).
CC OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC OC Streptococcus.
CC OX NCBI_TaxID=198466;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=MCAS315 / Serotype M3;
CC RX MEDLINE=22133808; PubMed=12122206;
CC RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
CC Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,
CC Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
CC Schlievert P.M., Musser J.M.;
CC "Genome sequence of a serotype M3 strain of group A Streptococcus:
CC RT phage-encoded toxins, the high-virulence phenotype, and clone
CC RT emergence."
CC RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=SSI-1 / Serotype M3;
CC RX MEDLINE=22683278; PubMed=12799345;
CC RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
CC Okashashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
CC Hayashi H., Hattori M., Hamada S.;
CC "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
CC RT large-scale genomic rearrangement in invasive strains and new insights
CC RT into phage evolution."
CC RL Genome Res. 13:1042-1055(2003).
CC CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC CC -!- SUBUNIT: Monomer (by similarity).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB014171; AR080416.1; -
EMBL; AP005146; BAC64902.1; -
HAWAP; MF_00123; -; 1.
InterPro; IPR001278; Arg_tRNA-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; tRNA-synt_id_C.
InterPro; IPR001412; tRNA-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; tRNA-synt_id; 1.
Pfam; PF05746; tRNA-synt_id_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMS; TIGR00456; args; 1.
PROSITE; PS00178; AA TRNA_LIGASE I; FALSE NEG.
Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
SITE 121 131 "HIGH" REGION.
SEQUENCE 563 AA; 63134 MW; BOB7DEC31A9DCF63 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
Best Local Similarity 45.5%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
Db 94 EQVITAGSDYA 104

RESULT 14
SYR_STRP8
ID SYR_STRP8 STANDARD; PRT; 563 AA.
AC Q8NZ22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
GN ARG8 OR SPYM8_2183.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S., Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.P., Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q., Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RA "Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4658-4673 (2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP + diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB010119; AAL98627.1; -
HAWAP; MF_00123; -; 1.

InterPro; IPR001278; Arg_tRNA-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; tRNA-synt_id_C.
InterPro; IPR001412; tRNA-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; tRNA-synt_id; 1.
Pfam; PF05746; tRNA-synt_id_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMS; TIGR00456; args; 1.
PROSITE; PS00178; AA TRNA_LIGASE I; FALSE NEG.
Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
SITE 121 131 "HIGH" REGION.
SEQUENCE 563 AA; 63134 MW; PECAF176A68DB85B CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
Best Local Similarity 45.5%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXDYS 11
Db 94 EQVITAGSDYA 104

RESULT 15
SYR_STRPY
ID SYR_STRPY STANDARD; PRT; 563 AA.
AC Q95XL5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
GN ARG8 OR SPY2151.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP + diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE006633; AAK34788.1; -
HAWAP; MF_00123; -; 1.

InterPro; IPR001278; Arg_tRNA-synt_lc.
InterPro; IPR005148; N.
InterPro; IPR008909; tRNA-synt_id_C.
InterPro; IPR001412; tRNA-synt_I.
Pfam; PF03485; N-Arg; 1.
Pfam; PF00750; tRNA-synt_id; 1.
Pfam; PF05746; tRNA-synt_id_C; 1.
PRINTS; PR01038; TRNASYNTHARG.
TIGRFAMS; TIGR00456; args; 1.

DR PROSITE; PS00178; AA TRNA LIGASE I; FALSE NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome. 131 "HIGH" REGION.
FT SITE 121
SQ SEQUENCE 563 AA; 63120 MW; E0FZCAC28D03B613 CRC64;
Query Match 66.0%; Score 33; DB 1; Length 563;
Best Local Similarity 45.5%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGDYS 11
Db 94 EQVITAGSDYA 104

Search completed: June 3, 2004, 11:49:56
Job time : 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 EEVVPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_cheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	74.0	149	16	Q82ZB4 enterococc
2	36	72.0	238	2	Q82367 rhizobium t
3	36	72.0	319	16	Q8ESV7
4	36	72.0	363	17	Q30260 archaeoglob
5	36	72.0	595	10	Q946J7 andrographi
6	36	72.0	1044	16	Q8DIH0 synchococc
7	35	70.0	78	6	Q9XST4 canis fami
8	35	70.0	175	6	Q81033 bos taurus
9	35	70.0	215	6	Q81031 bos taurus
10	35	70.0	217	4	Q00404 homo sapien
11	35	70.0	281	6	Q867A5 tragulus ja
12	35	70.0	290	16	Q8U7J0 agrobacteri
13	35	70.0	237	6	Q8HXI9 bos taurus
14	35	70.0	239	4	Q9UEE9 Queses homo sapien
15	35	70.0	587	16	Q9JZP8 Q9jzp8 neisseria m
16	35	70.0	692	5	Q9VI90 drosophila

ALIGNMENTS

RESULT 1

Q82ZB4 ID Q82ZB4 PRELIMINARY; PET; 149 AA.
AC Q82ZB4;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Large conductance mechanosensitive channel protein.
GN MSCL OR EF3152.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OK NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri R.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RA "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis";
RL Science 299:2071-2074(2003).
DR EMBL; AF016956; AAC02828.1; -.
DR TIGR; EF3152; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005216; F:ion channel activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001185; MS_channel.
DR Pfam; PF01741; Mscl; 1.
DR PRINTS; PR01264; MECHCHANNEL.
DR PRODOM; PD007253; MS_Channel; 1.
DR TIGRFAMs; TIGR00220; mscl; 1.
KW Complete proteome.
SQ SEQUENCE 149 AA; 16127 MW; 555799BF1E47D34E CRC64;

Query Match 74.0%; Score 37; DB 16; Length 149;

```

Best Local Similarity 70.0%; Pred. No. 5.2;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGXDY 10
Db 113 EVVVPXGXDY 122

RESULT 2
OS2367 PRELIMINARY; PRT; 298 AA.
AC OS2367;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aryl-alcohol dehydrogenase homolog (Fragment).
GN XYLBI.
OS Rhizobium tropici.
OG Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OC NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CFN299;
RA Rosenblueth M., Hynes M.F., Martinez-Romero E.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR ENBL; AF036920; AAC04779.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR Metal-binding; Oxidoreductase; Zinc; Plasmid.
FT NON TER 298
SQ SEQUENCE 298 AA; 31092 MW; 49B2F8117C33AE87 CRC64;

Query Match 72.0%; Score 36; DB 2; Length 298;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGXDYS 11
Db 250 EIPEGADFS 259

RESULT 3
Q8ESV7 PRELIMINARY; PRT; 319 AA.
AC Q8ESV7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical conserved protein.
GN OB0509.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OC NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."

```

```

RL Nucleic Acids Res. 30:3927-3935 (2002).
DR ENBL; AF004594; BAC12465.1; -.
DR InterPro; IPR001279; Bactnase-like.
DR Pfam; PF00753; lactamase_B; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDAE4BF13E79E37 CRC64;

Query Match 72.0%; Score 36; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVVPXGXDY 10
Db 189 EQLVPHGIDY 198

RESULT 4
O30260 PRELIMINARY; PRT; 363 AA.
AC O30260;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein AF2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OC NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049443; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
DR ENBL; AE001109; AAB91255.1; -.
DR PIR; D69551; D69551.
DR TIGR; AF2411; -.
DR InterPro; IPR002103; Bact_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;

Query Match 72.0%; Score 36; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGXDYS 11
Db 120 ENIVPYGIDFS 130

RESULT 5
Q946J7 PRELIMINARY; PRT; 595 AA.
AC Q946J7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (EC 1.1.1.34).
DE HMGR1.
GN

```

CC Androgaphis paniculata.
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
CC Lamiales; Lamiales; Acanthaceae; Acanthoideae; Ruscaceae;
CC Androgaphis paniculata.
CC NCBI_TaxID=175694;
RN [1]
RN SEQUENCE FROM N.A.
RA Krishnan S., Banerjee N.S.;
RT "3-hydroxy-3-methylglutaryl coenzyme A reductase gene from
RT Androgaphis paniculata."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF389879; AAL28015.2;
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0004420; F: hydroxymethylglutaryl-CoA reductase (NADPH) . . . ; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0009058; P: biosynthesis; IEA.
DR GO; GO:0006629; P: lipid metabolism; IEA.
DR InterPro; IPR009023; HMG-CoA_red.
DR InterPro; IPR009023; HMG-CoA_NAD_bind.
DR InterPro; IPR004554; HMG-CoA_R_NADP.
DR InterPro; IPR009029; HMG-CoA_sub_bind.
DR Pfam; PF00368; HMG-CoA_red; 1.
DR PRINTS; PR00071; HMGCOARCTASE.
DR TIGRFAMs; TIGR00533; HMG-CoA_R_NADP; 1.
DR PROSITE; PS00066; HMG-CoA_REDUCTASE_1; 1.
DR PROSITE; PS00318; HMG-CoA_REDUCTASE_2; 1.
DR PROSITE; PS01192; HMG-CoA_REDUCTASE_3; 1.
DR PROSITE; PS00065; HMG-CoA_REDUCTASE_4; 1.
DR Oxidoreductase.
KW Oxidoreductase.
SQ SEQUENCE 595 AA; 63268 MW; 19A3EA572F67AB2E CRC64;
Query Match 72.0%; Score 36; DB 10; Length 595;
Best Local Similarity 70.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVPXGXDYS 11
DB 15 EVAPPGHDYS 24
RESULT 6
Q8DIHO PRELIMINARY; PRT; 1044 AA.
AC Q8DIHO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RC MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Matsuura Y., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1";
RL DNA Res. 9:123-130(2002).
DR ENBL; AF005374; BAC09170.1;
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005215; F: transporter activity; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR001036; Actflvin_res.
DR InterPro; IPR004764; HAEI.
DR Pfam; PF00873; ACR_tran.1.
DR PRINTS; PR00702; ACRFLAVINRP.
DR TIGRFAMs; TIGR00915; 2A0602; 1.

KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
Query Match 72.0%; Score 36; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPXGXDYS 11
DB 843 EEVLPNGIGYS 853
RESULT 7
Q9XST4 PRELIMINARY; PRT; 78 AA.
AC Q9XST4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P97 homologous protein (Fragment).
GN P97.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=20422104; PubMed=10964405;
RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
RA Christophe D.;
RT "A method for the large-scale cloning of nuclear proteins and nuclear
RT targeting sequences on a functional basis."
RL Anal. Biochem. 284:231-239(2000).
DR EMBL; AJ388531; CAB46833.1; -;
FT NON TER 78
FT SEQUENCE 78 AA; 8895 MW; B62486313555FEAL CRC64;
Query Match 70.0%; Score 35; DB 6; Length 78;
Best Local Similarity 54.5%; Pred. No. 6.8;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPXGXDYS 11
DB 16 EDVVPSGEYS 26
RESULT 8
Q8I033 PRELIMINARY; PRT; 175 AA.
AC Q8I033;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Human-type bcl-2 protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Jersey;
RC Iwashita S., Itoh T.;
RT "A LINE-mediated gene diversity";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081003; BAC11952.1; -;
DR SEQUENCE 175 AA; 19529 MW; CE7283CB98393BB6 CRC64;
Query Match 70.0%; Score 35; DB 6; Length 175;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;


```

L Science 294:2323-2328(2001).
R EMBL; AE009374; AAL45253.1; -.
R PIR; AG3104; AG3104.
R PIR; D98182; D98182.
R GO; GO:0005222; C:intracellular; IEA.
R GO; GO:0003908; F:methylated-DNA-[protein]-cysteine S-methylt. . .; IEA.
R GO; GO:0008168; F:methyltransferase activity; IEA.
R GO; GO:0003700; F:transcription factor activity; IEA.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0006281; P:DNA repair; IEA.
R GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
R InterPro; IPR000005; HTHARAC.
R InterPro; IPR001497; Methyltransf_1.
R Pfam; PF00165; HTH_ARAC; 2.
R SMART; SM00342; Methyltransf_1.
R SMART; SM00342; HTH_ARAC; 1.
R TIGRFAMs; TIGR00589; Ogt; 1.
R PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
R Methyltransferase; Transferase; Complete proteome.
Q SEQUENCE 290 AA; 31587 MW; 5626592EF519977F CRC64;

Query Match 70.0%; Score 35; DB 16; Length 290;
Best Local Similarity 50.0%; Pred.No. 31;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Y 1 BEVVPXGXDY 10
b |::|||
9 EDITPIGSDY 18

RESULT 13
28HX9 PRELIMINARY; PRT; 297 AA.
ID Q8HX9
AC Q8HX9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Human-type Bcmt.
OS Bos taurus (Bovine).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
XC Bovidae; Bovinae; Bos.
X NCBI_TaxID=9913;
IN [1]
IP SEQUENCE FROM N.A.
IC TISSUE=Kidney;
CA Iwashita S., Itoh T.;
XT "A LINE-mediated gene diversity.";
XL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
JR EMBL; AB081004; BAC11953.1; -.
SQ SEQUENCE 297 AA; 33354 MW; D4A944BC8740373C CRC64;

Query Match 70.0%; Score 35; DB 6; Length 297;
Best Local Similarity 54.5%; Pred.No. 31;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Y 1 BEVVPXGXDYS 11
b |::|||
16 EDYVPSGGEYS 26

RESULT 14
29UEE9 PRELIMINARY; PRT; 299 AA.
ID Q9UEE9
AC Q9UEE9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BCNT protein (CRANIOFACIAL development protein 1).
GN BCNT.
OS Homo sapiens (Human).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98267221; PubMed=9602175;
RA Takahashi I., Ncbukuni T., Ohmori H., Kobayashi M., Tanaka S.,
RA Oshima K., Okada N., Masui T., Hashimoto K., Iwashita S.;
RT "Existence of a bovine LINE repetitive insert that appears in the cDNA
RT of bovine protein BCNT in ruminant, but not in human, genomes.";
RL Gene 211:387-394(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kalnine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB009285; BAA31867.1; -.
DR EMBL; BC000991; AAB00991.1; -.
DR EMBL; BT009819; AAP88821.1; -.
DR Genew; HGNC:1873; CFDP1.
SQ SEQUENCE 299 AA; 33593 MW; F4A9E28B669451A CRC64;

Query Match 70.0%; Score 35; DB 4; Length 299;
Best Local Similarity 54.5%; Pred.No. 32;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Y 1 BEVVPXGXDYS 11
b |::|||
16 EDYVPSGGEYS 26

RESULT 15
Q9JZP8 PRELIMINARY; PRT; 587 AA.
ID Q9JZP8
AC Q9JZP8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Succinate dehydrogenase, flavoprotein subunit.
GN NMB0950.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=2017575; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citton H., Clark B.B.,
RA Cotton M.D., Utterback T.R., Knouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE002446; AAF41356.1; -.
DR PIR; F81138; F81138.
DR HSP; P00363; IKF6.
DR TIGR; NMB0950; -.
DR GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR003953; FAD_bind2.
DR InterPro; IPR001327; FAD_Py_redox.

```

DR InterPro: IPR003952; FRD/SDH_FAD_BS.
DR InterPro: IPR001100; Pyr_redox.
DR InterPro: IPR004112; Succ_DH flav_C.
DR Pfam: PF00890; FAD_binding_2; 1.
DR Pfam: PF02910; succ_DH flav_C; 1.
DR PRINTS: PR00368; PADERN.
DR PRINTS: PR00411; ENDRITASEI.
DR PROSITE: PS00504; FRD_SDH_FAD_BINDING; 1.
KW FAD; Flavoprotein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 587 AA; 64502 MW; 9581701B08069003 CRC64;

Query Match 70.0%; Score 35; DB 16; Length 587;
Best Local Similarity 70.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGXDY 10
Db 366 EVVVPQGEDY 375

Search completed: June 3, 2004, 11:57:35
Job time : 30.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 3, 2004, 11:31:01 ; Search time 45.9333 Seconds
(without alignment)
67.664 Million cell updates/sec

Title: US-09-909-164-47

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseqp_29Jan04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	50	96.2	11	5	ABB80525	Hepatitis
2	50	96.2	11	5	ABB80521	Hepatitis
3	50	96.2	11	5	ABB80522	Hepatitis
4	50	96.2	11	5	ABB80566	Hepatitis
5	50	96.2	11	5	ABB80563	Hepatitis
6	50	96.2	11	5	ABB80565	Hepatitis
7	50	96.2	11	5	ABB80567	Hepatitis
8	50	96.2	11	5	ABB80559	Hepatitis
9	50	96.2	11	5	ABB80526	Hepatitis
10	50	96.2	11	5	ABB80564	Hepatitis
11	50	96.2	11	5	ABB80568	Hepatitis
12	46	88.5	11	5	ABB80561	Hepatitis
13	46	88.5	11	5	ABB80524	Hepatitis
14	46	88.5	11	5	ABB80529	Hepatitis
15	46	88.5	11	5	ABB80528	Hepatitis
16	46	88.5	11	5	ABB80562	Hepatitis
17	45	86.5	11	5	ABB80523	Hepatitis
18	45	86.5	11	5	ABB80536	Hepatitis
19	45	86.5	11	5	ABB80558	Hepatitis
20	45	86.5	11	5	ABB80560	Hepatitis
21	45	86.5	11	5	ABB80527	Hepatitis
22	45	86.5	11	5	ABB80535	Hepatitis
23	45	86.5	11	5	ABB80540	Hepatitis
24	45	86.5	11	5	ABB80539	Hepatitis
25	44	84.6	11	5	ABB80549	Hepatitis

26	44	84.6	11	5	ABB80544	Hepatitis
27	44	84.6	11	5	ABB80553	Hepatitis
28	44	84.6	11	5	ABB80552	Hepatitis
29	44	84.6	11	5	ABB80545	Hepatitis
30	42	80.8	11	5	ABB80530	Hepatitis
31	41	78.8	11	5	ABB80542	Hepatitis
32	41	78.8	11	5	ABB80543	Hepatitis
33	41	78.8	11	5	ABB80538	Hepatitis
34	40	76.9	11	5	ABB80548	Hepatitis
35	40	76.9	11	5	ABB80547	Hepatitis
36	40	76.9	11	5	ABB80556	Hepatitis
37	40	76.9	11	5	ABB80557	Hepatitis
38	40	76.9	11	5	ABB80537	Hepatitis
39	40	76.9	11	5	ABB80551	Hepatitis
40	40	76.9	11	5	ABB80541	Hepatitis
41	40	76.9	20	2	AAU76810	Novel hum
42	40	76.9	1022	4	ABG03621	Novel hum
43	40	76.9	1022	4	ABG08173	Novel hum
44	40	76.9	1022	4	ABG05826	Novel hum
45	39	75.0	11	5	ABB80546	Hepatitis

ALIGNMENTS

RESULT 1
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX

AC ABB80525;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX
OS Synthetic.

PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WC-US023169.

21-JUL-2000; 2000US-0220101P.

(CORV-) CORVAS INT INC.

Lim-Wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease.

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 2
ABB80521
ID ABB80521 standard; peptide; 11 AA.
XX
AC ABB80521;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX
XX Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 2
ABB80521
ID ABB80521 standard; peptide; 11 AA.
XX
AC ABB80521;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX
XX Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 3
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
AC ABB80522;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX
XX Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

RESULT 4
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
AC ABB80522;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6
FT Misc-difference 9 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C virus
FT protease.
XX
XX Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMSYS 11

AB80566	D	AB80566 standard; peptide; 11 AA.
XX	CC	AB80566;
XX	CC	08-OCT-2002 (first entry)
XX	CC	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX	CC	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
XX	CC	Synthetic.
XX	CC	Key
XX	CC	Location/Qualifiers
XX	CC	Modified-site 1
XX	CC	/note= "N-terminal acetyl"
XX	CC	Modified-site 6
XX	CC	/note= "2-aminoisobutyryl carbonyl residue forming a keto
XX	CC	-amide linkage with residue 7"
XX	CC	Modified-site 11
XX	CC	/note= "C-terminal amide"
XX	CC	WO200208251-A2.
XX	CC	31-JAN-2002.
XX	CC	19-JUL-2001; 2001WO-US023169.
XX	CC	21-JUL-2000; 2000US-0220101P.
XX	CC	(CORV-) CORVAS INT INC.
XX	CC	Lim-Wilby M, Levy OE, Brunck TK;
XX	CC	WPI; 2002-361643/39.
XX	CC	Novel peptide compound having hepatitis C virus protease inhibitory
XX	CC	activity useful for treating disorders associated with hepatitis C virus
XX	CC	proteases.
XX	CC	Claim 17; Page 65; 69pp; English.
XX	CC	The sequence represents a peptide compound of the invention having
XX	CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX	CC	invention are alpha-ketoamide peptide analogues. The peptides have
XX	CC	virucide activity, and are useful for treating and in the manufacture of
XX	CC	a medicament to treat disorders associated with HCV protease. A
XX	CC	pharmaceutical composition comprising the peptide as an active ingredient
XX	CC	is useful for treating disorders associated with hepatitis C virus
XX	CC	Sequence 11 AA;
XX	CC	Query Match 96.2%; Score 50; DB 5; Length 11;
XX	CC	Best Local Similarity 100.0%; Pred. No. 0.002;
XX	CC	Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX	CC	QY 1 EEVVPXGMSYS 11
XX	CC	
XX	CC	Db 1 EEVVPXGMSYS 11
XX	CC	
XX	CC	RESULT 5
XX	CC	AB80563
XX	CC	ID AB80563 standard; peptide; 11 AA.
XX	CC	AC AB80563;
XX	CC	XX
XX	CC	DT 08-OCT-2002 (first entry)
XX	CC	XX
XX	CC	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.
XX	CC	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

```

KW virucide.
XX
XX Synthetic.
XX
XX Location/Qualifiers
FH 1
FT Modified-site 1
FT FT /note= "N-terminal acetyl"
FT Modified-site 6
FT FT /note= "Valyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT FT /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US023169.
XX
XX 21-JUL-2000; 2000US-0220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-Wilby M, Levy OB, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus
XX
XX Sequence 11 AA;
XX
XX Query Match 96.2%; Score 50; DB 5; Length 11;
XX Best Local Similarity 100.0%; Pred. NO. 0.002;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
QY 1 EEVVPXGMSYS 11
Db 11
1 EEVVPXGMSYS 11
XX
RESULT 6
ABB80565
ID ABB80565 standard; peptide; 11 AA.
XX
AC ABB80565;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH 1
FT Modified-site 1
FT FT /note= "N-terminal acetyl"
FT Modified-site 6
FT FT /note= "Norleucyl carbonyl forming keto-amide linkage
FT FT with residue 7"

```


XX PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C virus
PT protease.
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 9
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX AC ABB80526;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
PT protease.
XX PT activity useful for treating disorders associated with hepatitis C virus
XX protease.
XX Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 10
ABB80564
ID ABB80564 standard; peptide; 11 AA.
XX AC ABB80564;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Leucyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11
RESULT 10
ABB80564
ID ABB80564 standard; peptide; 11 AA.
XX AC ABB80564;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Leucyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US023169.
XX PR 21-JUL-2000; 2000US-0220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-Wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus
XX SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002; 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
 DB 1 BEVVPXGMSYS 11

RESULT 11
 ABB80568
 ID ABB80568 standard; peptide; 11 AA.
 XX AC ABB80568;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX OS virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
 FT a keto-amide linkage with residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 XX PF 19-JUL-2001; 2001WO-US023169.
 XX PR 21-JUL-2000; 2000US-0220101P.
 XX PA (CORV-) CORVAS INT INC.
 XX PI Lim-Wilby M, Levy OE, Brunck TK;
 XX DR WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitor
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX PS Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Query Match 96.2%; Score 50; DB 5; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
 DB 1 BEVVPXGMSYS 11

RESULT 12
 ABB80561
 ID ABB80561 standard; peptide; 11 AA.
 XX AC ABB80561;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX OS virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 XX PF 19-JUL-2001; 2001WO-US023169.
 XX PR 21-JUL-2000; 2000US-0220101P.
 XX PA (CORV-) CORVAS INT INC.
 XX PI Lim-Wilby M, Levy OE, Brunck TK;
 XX DR WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C virus
 PT protease.
 XX PS Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus

Query Match 88.5%; Score 46; DB 5; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.013;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
 DB 1 BEVVPXGMSYS 11

RESULT 13
 ABB80524
 ID ABB80524 standard; peptide; 11 AA.
 XX AC ABB80524;
 XX

08-OCT-2002 (first entry)
Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
Synthetic.
Key Location/Qualifiers
Modified-site 1 /note= "N-terminal acetyl"
Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"
Misc-difference 8 /note= "D-form residue"
Misc-difference 9 /note= "D-form residue"
Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C virus
protease.
Claim 17; Page 64; 69pp; English.
The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus
protease.
Sequence 11 AA;
Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMDYS 11
RESULT 14
ABB0529
ID ABB0529 standard; peptide; 11 AA.
AC ABB0529;
XX 08-OCT-2002 (first entry)
DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.

08-OCT-2002 (first entry)
Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
Synthetic.
Key Location/Qualifiers
Modified-site 1 /note= "N-terminal acetyl"
Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"
Misc-difference 9 /note= "D-form residue"
Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C virus
protease.
Claim 17; Page 64; 69pp; English.
The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus
protease.
Sequence 11 AA;
Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMDYS 11
RESULT 15
ABB0528
ID ABB0528 standard; peptide; 11 AA.
AC ABB0528;
XX 08-OCT-2002 (first entry)
DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
WO200208251-A2.
31-JAN-2002.
19-JUL-2001; 2001WO-US023169.
21-JUL-2000; 2000US-0220101P.
(CORV-) CORVAS INT INC.
Lim-Wilby M, Levy OE, Brunck TK;
WPI; 2002-361643/39.
Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C virus
protease.
Claim 17; Page 64; 69pp; English.
The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus
protease.
Sequence 11 AA;
Query Match 88.5%; Score 46; DB 5; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMDYS 11
RESULT 15
ABB0528
ID ABB0528 standard; peptide; 11 AA.
AC ABB0528;
XX 08-OCT-2002 (first entry)
DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX Synthetic.
Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

Fri Jun 4 16:16:15 2004

```
FT      residue 7"
FT      Misc-difference 8
FT      /note= "D-form residue"
FT      Modified-site 11
FT      /note= "C-terminal amide"
XX
XX      WO200208251-A2.
XX
XX      31-JAN-2002.
XX
XX      19-JUL-2001; 2001WO-US023169.
XX
XX      21-JUL-2000; 2000US-0220101P.
XX
XX      (CORV-) CORVAS INT INC.
XX
XX      Lim-Wilby M, Levy OE, Brunck TK;
XX      WPI; 2002-361643/39.
XX
XX      Novel peptide compound having hepatitis C virus protease inhibitory
FT      activity useful for treating disorders associated with hepatitis C virus
FT      protease.
XX
XX      Claim 17; Page 64; 69pp; English.
XX
XX      The sequence represents a peptide compound of the invention having
CC      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC      invention are alpha-ketoamide peptide analogues. The peptides have
CC      virucide activity, and are useful for treating and in the manufacture of
CC      a medicament to treat disorders associated with HCV protease. A
CC      pharmaceutical composition comprising the peptide as an active ingredient
CC      is useful for treating disorders associated with hepatitis C virus
XX
XX      SQ      Sequence 11 AA;
XX
XX      Query Match      88.5%; Score 46; DB 5; Length 11;
XX      Best Local Similarity 90.9%; Pred. No. 0.013;
XX      Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX      QY      1 EEVVPXGMSYS 11
XX      Db      1 EEVVPXGMDYS 11
XX
XX      Search completed: June 3, 2004, 11:48:25
XX      Job time : 45.9333 secs
```



```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (327)..(328)
; OTHER INFORMATION: Amino acids 327 & 328 are Xaa wherein Xaa = any amino acid.
US-09-134-000C-3738
Query Match      69.2%; Score 36; DB 4; Length 382;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
   :| |||||
Db 332 LIEPGMSYS 340

RESULT 3
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 635918
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match      65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
   :| |||||
Db 686 VMPGMSYS 694

RESULT 4
US-09-540-236-2902
; Sequence 2902, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2902
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2902

Query Match      65.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSY 10
   :||| |||
Db 783 EILPVGWY 791

RESULT 5
US-09-328-352-7885
; Sequence 7885, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match      65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSY 10
   :||| |||
Db 596 EVVPEGLSP 604

RESULT 6
US-09-760-946-2
; Sequence 2, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghio, Elise
; APPLICANT: Goudreau, Nathalie
; APPLICANT: Halmos, Teddy
; APPLICANT: Liinas-Brunet, Montse
; TITLE OF INVENTION: Macrocytic Peptides Active Against the Hepatitis C Virus
; FILE REFERENCE: 13/076-1-C1
; CURRENT APPLICATION NUMBER: US/09/760,946
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/542,675
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/128,011
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Substrate for recombinant HCV NS3 protease radiometric assay
US-09-760-946-2

Query Match      63.5%; Score 33; DB 4; Length 12;
Best Local Similarity 45.5%; Pred. No. 1.7;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVPXGMSYS 11
   :||| |||
Db 1 DDIVPCMSYT 11

RESULT 7
US-09-760-946-3
; Sequence 3, Application US/09760946
; Patent No. 6608027
; GENERAL INFORMATION:
; APPLICANT: Tsantrizos, Youla S.
; APPLICANT: Cameron, Dale R.
; APPLICANT: Faucher, Anne-Marie
; APPLICANT: Ghio, Elise

```

```

; APPLICANT: Goudreau, Nathalie
; APPLICANT: Halmos, Teddy
; APPLICANT: Llinas-Brunet, Montse
; TITLE OF INVENTION: Macrocyclic Peptides Active Against the Hepatitis C Virus
; FILE REFERENCE: 13/076-1-C1
; CURRENT APPLICATION NUMBER: US 09/760,946
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/542,675
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/128,011
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tracer for NS3 protease assay
; NAME/KEY: MOD_RES
; LOCATION: (1)_
; OTHER INFORMATION: Asp at position 1 is biotinylated
; NAME/KEY: MOD_RES
; LOCATION: (10)
; OTHER INFORMATION: Tyr at position 10 is iodinated with I-125
US-09-760-946-3

```

Query Match 63.5%; Score 33; DB 4; Length 12;
 Best Local Similarity 45.5%; Pred. No. 1.7; Indels 2; Gaps 0;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
 Db 1 DBIVPCSMYS 11

```

RESULT 8
US-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637.759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RIMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-637-759B-236

```

Query Match 63.5%; Score 33; DB 2; Length 45;
 Best Local Similarity 60.0%; Pred. No. 7.8; Indels 3; Gaps 0;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
 Db 1 BEISPLGWSY 10

```

RESULT 9
US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RIMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-871-355A-236

```

Query Match 63.5%; Score 33; DB 3; Length 45;
 Best Local Similarity 60.0%; Pred. No. 7.8; Indels 3; Gaps 0;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
 Db 1 BEISPLGWSY 10

```
RESULT 10
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 1 BEISPLQWSY 10

RESULT 11
5177197-51
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:51:
; LENGTH: 65
5177197-51
Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
Db 399 KEICPGMGYT 409

RESULT 12
5177197-1
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:1:
; LENGTH: 410
5177197-1
Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 97;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
Db 399 KEICPGMGYT 409

RESULT 13
5177197-30
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:30:
; LENGTH: 1394
5177197-30
Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
Db 399 KEICPGMGYT 409

RESULT 14
US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whole
; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Protease
; TITLE OF INVENTION: Other Enzymes and the Use Thereof
; FILE REFERENCE: 1735.0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642
```

EARLIER FILING DATE: 21-JUL-1998
NUMBER OF SEQ ID NOS: 139
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: Peptide
JS-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

2y 1 BEVVPXGMSY 10
:::|
db 1 DDIVPCMSY 10

RESULT 15

JS-09-521-650-66
Sequence 66, Application US/09521650
Patent No. 6335429
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.
APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
TITLE OF INVENTION: Use Thereof
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521,650
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: Peptide
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred.No. 2.3;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

2y 1 BEVVPXGMSY 10
:::|
db 1 DDIVPCMSY 10

Search completed: June 3, 2004, 12:03:09
Job time : 11.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M protein - protein search, using sw model

Run on: June 3, 2004, 11:57:42 ; Search time 33.7333 Seconds
(without alignments)
91.741 Million cell updates/sec

Title: US-09-909-164-47

Perfect score: 52

Sequence: 1 EHVVPKMSYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	12	US-09-909-164-5
2	50	96.2	11	12	US-09-909-164-6
3	50	96.2	11	12	US-09-909-164-9
4	50	96.2	11	12	US-09-909-164-10
5	50	96.2	11	12	US-09-909-164-47
6	50	96.2	11	12	US-09-909-164-48
7	50	96.2	11	12	US-09-909-164-49
8	50	96.2	11	12	US-09-909-164-50
9	50	96.2	11	12	US-09-909-164-51
10	50	96.2	11	12	US-09-909-164-52
11	46	88.5	11	12	US-09-909-164-8
12	46	88.5	11	12	US-09-909-164-12
13	46	88.5	11	12	US-09-909-164-13
14	46	88.5	11	12	US-09-909-164-7
15	45	86.5	11	12	US-09-909-164-11

16	45	86.5	11	12	US-09-909-164-19
17	45	86.5	11	12	US-09-909-164-20
18	45	86.5	11	12	US-09-909-164-23
19	45	86.5	11	12	US-09-909-164-24
20	44	84.6	11	12	US-09-909-164-28
21	44	84.6	11	12	US-09-909-164-29
22	44	84.6	11	12	US-09-909-164-33
23	44	84.6	11	12	US-09-909-164-36
24	44	84.6	11	12	US-09-909-164-37
25	44	84.6	11	12	US-09-909-164-43
26	42	80.8	11	12	US-09-909-164-14
27	41	78.8	11	12	US-09-909-164-22
28	41	78.8	11	12	US-09-909-164-26
29	41	78.8	11	12	US-09-909-164-27
30	41	78.8	11	12	US-09-909-164-61
31	41	78.8	11	12	US-09-909-164-62
32	40	76.9	11	12	US-09-909-164-21
33	40	76.9	11	12	US-09-909-164-25
34	40	76.9	11	12	US-09-909-164-31
35	40	76.9	11	12	US-09-909-164-32
36	40	76.9	11	12	US-09-909-164-35
37	40	76.9	11	12	US-09-909-164-40
38	40	76.9	11	12	US-09-909-164-41
39	40	76.9	11	12	US-09-909-164-45
40	40	76.9	11	12	US-09-909-164-46
41	39	75.0	11	12	US-09-909-164-30
42	39	75.0	11	12	US-09-909-164-34
43	39	75.0	11	12	US-09-909-164-38
44	39	75.0	11	12	US-09-909-164-39
45	39	75.0	11	12	US-09-909-164-42

ALIGNMENTS

RESULT 1

US-09-909-164-5
; Sequence 5, Application US/0909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN0192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;

```
Best Local Similarity 100.0%; Pred. No. 0.0014; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;

QY 1 EVVVPXGMSYS 11
   |||||
Db 1 EVVVPXGMSYS 11

RESULT 2
US-09-909-164-6
; Sequence 6, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-9

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
   |||||
Db 1 EVVVPXGMSYS 11

RESULT 4
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
```

```
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
; US-09-909-164-9

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMSYS 11
   |||||
Db 1 EVVVPXGMSYS 11

RESULT 4
US-09-909-164-10
; Sequence 10, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunnck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
```

OTHER INFORMATION: D-amino acids

JS-09-909-164-10

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 BEVVPXGMSYS 11
| | | | | | | | | | |
b 1 BEVVPXGMSYS 11

RESULT 5

JS-09-909-164-47
Sequence 47, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: valine-(CO)
JS-09-909-164-47

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 BEVVPXGMSYS 11
| | | | | | | | | | |
b 1 BEVVPXGMSYS 11

RESULT 6

JS-09-909-164-48
Sequence 48, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: leucine-(CO)
US-09-909-164-48

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 BEVVPXGMSYS 11
| | | | | | | | | | |
b 1 BEVVPXGMSYS 11

RESULT 7

US-09-909-164-49
Sequence 49, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norleucine-(CO)
US-09-909-164-49

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 BEVVPXGMSYS 11

Db 1 EVVFXGMSYS 11
|||||

RESULT 8

US-09-909-164-50
; Sequence 50, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: 2-amino-butyric acid- (CO)
US-09-909-164-50

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVFXGMSYS 11
|||||

Db 1 EVVFXGMSYS 11
|||||

RESULT 9

US-09-909-164-51
; Sequence 51, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: (s,s)-allothreonine- (CO)
US-09-909-164-51

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVFXGMSYS 11
|||||

Db 1 EVVFXGMSYS 11
|||||

RESULT 10

US-09-909-164-52
; Sequence 52, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: propynyl glycine- (CO)
US-09-909-164-52

Query Match 96.2%; Score 50; DB 12; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EVVFXGMSYS 11
|||||

Db 1 EVVFXGMSYS 11
|||||

RESULT 11

US-09-909-164-8
; Sequence 8, Application US/09909164
; Publication No. US20020068702A1


```
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
FILE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
FEATURE:
NAME/KEY: MOD RES
LOCATION: (1)..(11)
OTHER INFORMATION: ACETYLTATION
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
US-09-909-164-8
Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2y 1 EVVVPXGMSYS 11
   |||||
3b 1 EVVVPXGMDYS 11
   |||||

RESULT 12
US-09-909-164-12
; Sequence 12, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-13
; Sequence 13, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; OTHER INFORMATION: 11-mer synthesized according to example 1
US-09-909-164-13
Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
   |||||
Db 1 EVVVPXGMDYS 11
   |||||

RESULT 13
US-09-909-164-13
; Sequence 13, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(11)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; OTHER INFORMATION: 11-mer synthesized according to example 1
US-09-909-164-13
Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EVVVPXGMSYS 11
   |||||
Db 1 EVVVPXGMDYS 11
   |||||
```

RESULT 14
US-09-909-164-7
; Sequence 7, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
US-09-909-164-7

Query Match 86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMHYS 11

RESULT 15
US-09-909-164-11
; Sequence 11, Application US/09909164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Brunk, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: 11-mer synthesized according to example 1

; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline-(CO)
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
US-09-909-164-11
Query Match 86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 EEVVPXGMSYS 11
| | | | | | | | | |
Db 1 EEVVPXGMHYS 11
Search completed: June 3, 2004, 12:57:17
Job time : 33.7333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-47

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2 T31308	hypothetical 367K
2	37	71.2	840	2 T39116	probable sulfate p
3	37	71.2	877	2 T40413	sulfate permease -
4	36	69.2	102	2 A42452	V1 protein - tobac
5	36	69.2	1498	2 B97355	DNA segregation AT
6	35	67.3	225	2 S57810	hypothetical prote
7	35	67.3	425	2 T24111	hypothetical prote
8	35	67.3	670	2 S22293	zinc finger protei
9	35	67.3	749	2 H82691	topoisomerase IV s
10	35	67.3	217	2 A34203	DNA-binding protei
11	34	65.4	156	2 S54619	hypothetical prote
12	34	65.4	252	2 H69491	cell division inh
13	34	65.4	544	2 C82900	probable ABC sub
14	33	63.5	94	2 I40758	hypothetical prote
15	33	63.5	116	2 E90544	50S ribosomal prot
16	33	63.5	165	2 D69493	hypothetical prote
17	33	63.5	253	2 C81374	hypothetical prote
18	33	63.5	259	2 T34536	hypothetical prote
19	33	63.5	284	2 S75817	beta-ketacyl-ACP
20	33	63.5	298	2 T47670	hypothetical prote
21	33	63.5	368	2 F72281	hypothetical prote
22	33	63.5	426	2 D82163	3-phosphoshikimate
23	33	63.5	466	2 T43653	cdc37 protein - fi
24	33	63.5	653	2 D82352	iron(III) ABC tran
25	33	63.5	890	2 A30481	bacteriocin BCN5 -
26	33	63.5	1028	2 AF3286	ATP-dependent DNA
27	33	63.5	1152	2 D87046	conserved hypothet
28	33	63.5	1394	2 A35626	transforming growt
29	33	63.5	1401	2 G82336	DNA-directed RNA p

30 33 63.5 1548 2 T04456
31 33 63.5 1712 2 A38261
32 32 61.5 84 2 B97333
33 32 61.5 175 2 PQ0616
34 32 61.5 223 2 T01457
35 32 61.5 279 2 B72481
36 32 61.5 288 2 JC4011
37 32 61.5 288 2 I58372
38 32 61.5 289 2 A41984
39 32 61.5 289 2 A42822
40 32 61.5 291 2 S57922
41 32 61.5 291 2 S57925
42 32 61.5 291 2 JC4579
43 32 61.5 291 2 S62730
44 32 61.5 292 2 B42822
45 32 61.5 295 2 A38977

ALIGNMENTS

RESULT 1

T31308
Hypothetical 367K protein - Cenarchaeum symbiosum
C:Species: Cenarchaeum symbiosum
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T31308
R:Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180: 5003-5009, 1998
A:Title: Genomic analysis reveals chromosomal variation in natural populations of the u
A:Reference number: Z20994; MUID:98422450; PMID:9748430
A:Accession: T31308
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3472 <SCH>
A:Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC62699.1
C:Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. No. 60;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11

Db 2294 EDVTPRGISFS 2304

RESULT 2

T39116
Probable sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39116
R:Hunt, C.; Aves, S.; Mcbougall, R.C.; Rajandream, M.A.; Barrell, B.G.
Submitted to the EMBL Data Library, November 1999
A:Reference number: Z21829
A:Accession: T39116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-840 <HUN>
A:Cross-references: EMBL:AL132779; PIDN:CAB60015.1; GSPDB:GN000066; SPDB:SPAC869.05C
A:Experimental source: strain 972h-; cosmid c869
C:Genetics:
A:Gene: SPDB:SPAC869.05C
A:Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

Db 135 VVPQGNSTYA 143

RESULT 3
40413
ulfate permease - fission yeast (Schizosaccharomyces pombe)
;Species: Schizosaccharomyces pombe
;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
;Accession: T40413
;Lynne, M.; Rejandram, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
;Reference number: Z21926
;Accession: T40413
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-877 <LYN>
;Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
;Experimental source: strain 972h-; cosmid c3H7
;Genetics:
;Gene: SPDB:SPBC3H7.02
;Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2y 3 EVVPGMSYS 11
||| |||
Db 148 VVPGMSYA 156

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
;Species: tobacco yellow dwarf virus
;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
;Accession: A42452
;Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
;Virology 187, 633-642, 1992
;Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
;Reference number: A42452; UID:9188536; PMID:1546458
;Accession: A42452
;Molecule type: DNA
;Residues: 1-102 <WOR>
;Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPGMSYS 11
: ||| |||
Db 7 QVPGSINYS 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/spoIIIE family, YUKA B. subtilis ortholog [imported] - Clos
;Species: Clostridium acetobutylicum
;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
;Accession: B97355
;Nalling, J.; Sretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daily, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
; J. Bacteriol. 183, 4823-4838, 2001
;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
;Reference number: A96900; UID:21359325; PMID:21359325
;Accession: B97355
;Status: preliminary
;Molecule type: DNA
;Residues: 1-1498 <KUR>
;Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026814; GSPDB:GN00168
;Experimental source: Clostridium acetobutylicum ATCC84
;Genetics:
;Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
: : ||| |||
Db 1276 EQKIPMGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone tpp11) - tomato
;Species: Lycopersicon esculentum (tomato)
;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
;Accession: S57810
;Milligan, S.B.; Gasser, C.S.
; Plant Mol. Biol. 28, 691-711, 1995
;Title: Nature and regulation of pistil-expressed genes in tomato.
;Reference number: S57808; UID:95375233; PMID:7647301
;Accession: S57810
;Status: preliminary; nucleic acid sequence not shown
;Molecule type: mRNA
;Residues: 1-225 <MIL>
;Cross-references: EMBL:U20592; NID:G924625; PIDN:AAA80497.1; PID:G924626
;Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 11
: ||| |||
Db 32 DEVVPNGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
;Accession: T24111
;Percy, C.
submitted to the EMBL Data Library, October 1996
;Reference number: Z19842
;Accession: T24111
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-425 <WIL>
;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
;Experimental source: clone R10D12
;Genetics:
;Gene: CESP:R10D12.10
;Map position: 5
;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
: : ||| |||
Db 335 EQIVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
;Species: Rattus norvegicus (Norway rat)
;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
;Accession: S22293; 178656
;Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

RESULT 14
IS-09-909-164-7
Sequence 7, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
FILE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
TITLE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: D-amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
IS-09-909-164-7

Query Match 86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMHYS 11

RESULT 15
IS-09-909-164-11
Sequence 11, Application US/09909164
Publication No. US20020068702A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Lim-Wilby, Marguerita
APPLICANT: Levy, Odile E
APPLICANT: Bruck, Terence K
FILE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
TITLE REFERENCE: IN01192-US
CURRENT APPLICATION NUMBER: US/09/909,164
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: 60/220,101
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 11
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: 11-mer synthesized according to example 1

FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)..(1)
OTHER INFORMATION: ACETYLATION
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (11)..(11)
OTHER INFORMATION: AMIDATION
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (6)..(6)
OTHER INFORMATION: norvaline- (CO)
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (8)..(8)
OTHER INFORMATION: D-amino acid
US-09-909-164-11

Query Match 86.5%; Score 45; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 1 EEVVPXGMHYS 11

Search completed: June 3, 2004, 12:57:17
Job time : 33.7333 secs

```
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acid
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; US-09-909-164-8

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMDYS 11

RESULT 12
US-09-909-164-12
; Sequence 12, Application US/09/909,164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: D-amino acids
; US-09-909-164-13

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMDYS 11

RESULT 13
US-09-909-164-13
; Sequence 13, Application US/09/909,164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; US-09-909-164-13

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMDYS 11
```

```
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(8)
; OTHER INFORMATION: D-amino acid
; US-09-909-164-12

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMDYS 11

RESULT 13
US-09-909-164-13
; Sequence 13, Application US/09/909,164
; Publication No. US20020068702A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Lim-Wilby, Marguerita
; APPLICANT: Levy, Odile E
; APPLICANT: Bruck, Terence K
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS-3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01192-US
; CURRENT APPLICATION NUMBER: US/09/909,164
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: 60/220,101
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 11
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: 11-mer synthesized according to example 1
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLTATION
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (11)..(11)
; OTHER INFORMATION: AMIDATION
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: norvaline- (CO)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)..(9)
; OTHER INFORMATION: D-amino acids
; US-09-909-164-13

Query Match      88.5%; Score 46; DB 12; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0091;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMDYS 11
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:47 ; Search time 9 Seconds
(without alignments)
117.567 Million cell updates/sec

Title: US-09-909-164-47

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	T31308	hypothetical 367K
2	37	71.2	840	T39116	probable sulfate p
3	37	71.2	877	T40413	sulfate permease -
4	36	69.2	102	A42452	V1 protein - tobac
5	36	69.2	1498	B97355	DNA segregation AT
6	35	67.3	225	S57810	hypothetical prote
7	35	67.3	425	T24111	hypothetical prote
8	35	67.3	670	S22293	zinc finger protei
9	35	67.3	749	H82691	topoisomerase IV s
10	35	67.3	2717	A34203	DNA-binding protei
11	34	65.4	156	S54619	hypothetical prote
12	34	65.4	252	H69431	cell division inhi
13	34	65.4	544	K82900	probable ABC subst
14	33	63.5	94	T40758	hypothetical prote
15	33	63.5	116	E90544	50S ribosomal prot
16	33	63.5	165	D69493	hypothetical prote
17	33	63.5	253	C81374	hypothetical prote
18	33	63.5	259	T34536	hypothetical prote
19	33	63.5	284	S75817	hypothetical prote
20	33	63.5	298	T47670	beta-ketoacyl-ACP
21	33	63.5	368	F72281	hypothetical prote
22	33	63.5	426	D82163	3-phosphoshikimate
23	33	63.5	466	T43653	cdc37 protein - fi
24	33	63.5	653	D82352	iron(III) ABC tran
25	33	63.5	890	A30481	bacteriocin BCN5 -
26	33	63.5	1028	AF3286	ATP-dependent DNA
27	33	63.5	1152	D87046	conserved hypothet
28	33	63.5	1394	A35626	transforming growt
29	33	63.5	1401	G82336	DNA-directed RNA p

30	63.5	1548	2	T04456	hypothetical prote
31	63.5	1712	2	A38261	masking protein pr
32	61.5	84	2	E97333	hypothetical prote
33	61.5	175	2	PQ0616	transport protein
34	61.5	223	2	T01457	rho protein GDP-di
35	61.5	279	2	B72481	hypothetical prote
36	61.5	288	2	JC4011	cyclin D2 - rat
37	61.5	288	2	I58372	cyclin D2 - rat
38	61.5	289	2	A41984	cyclin D2 - mouse
39	61.5	289	2	A42822	cyclin D2 - human
40	61.5	291	2	S57922	cyclin D1 - Africa
41	61.5	291	2	S57925	cyclin D2 - Africa
42	61.5	291	2	JC4579	cyclin D2 - chicke
43	61.5	291	2	S62730	cyclin D1 - zebra
44	61.5	292	2	B42822	cyclin D3 - human
45	61.5	295	2	A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308

hypothetical 367K protein - Cenarchaeum symbiosum

C:Species: Cenarchaeum symbiosum

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000

C:Accession: T31308

R:Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.

J. Bacteriol. 180, 5003-5009, 1998

A:Title: Genomic analysis reveals chromosomal variation in natural populations of the ur

A:Reference number: Z20994; MUID:98422450; PMID:9748430

A:Accession: T31308

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3472 <SCH>

A:Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599394; PIDN:AAC62699.1

C:Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;

Best Local Similarity 54.5%; Pred. No. 60;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11

Db 2294 EDVIPRGISFS 2304

RESULT 2

T39116

probable sulfate permease - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39116

R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, November 1999

A:Reference number: Z21829

A:Accession: T39116

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-840 <HUN>

A:Cross-references: EMBL:AL132779; PIDN:CAB60015.1; GSPDB:GN00066; SPDB:SPAC869.05C

A:Experimental source: strain 972h-; cosmid c869

C:Genetics:

A:Gene: SPDB:SPAC869.05C

A:Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;

Best Local Similarity 77.8%; Pred. No. 21;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11

Db 135 VVPXGMSYA 143

RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40413
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z21926
A;Accession: T40413
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-877 <LYN>
A;Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBCH7.02
A;Experimental source: strain 972h; cosmid C3H7
C;Genetics:
A;Gene: SPDB:SPBCH7.02
A;Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| |||
Db 148 VVPQMSYA 156

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C;Species: tobacco yellow dwarf virus
C;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C;Accession: A42452
R;Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A;Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A;Reference number: A42452; MUID:92188538; PMID:1546458
A;Accession: A42452
A;Molecule type: DNA
A;Residues: 1-102 <MOR>
A;Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
||| |||
Db 7 QVVPFGINYS 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clo
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: B97355
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97355
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1498 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026814; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||| |||
Db 1276 EQIPMGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPPI1) - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C;Accession: S57810
R;Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A;Title: Nature and regulation of pistil-expressed genes in tomato.
A;Reference number: S57808; MUID:95375233; PMID:7647301
A;Accession: S57810
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-225 <MIL>
A;Cross-references: EMBL:U20592; NID:924625; PIDN:AAA80497.1; PID:G924626
C;Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
||| |||
Db 32 DEVVPGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111
R;Percy, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19842
A;Accession: T24111
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z81109; PIDN:CA803241.1; GSPDB:GN00023; CESP:R10D12.10
A;Experimental source: clone R10D12
C;Genetics:
A;Gene: CESP:R10D12.10
A;Map position: 5
A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||| |||
Db 335 EQIVFGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C;Accession: S22293; 178656
R;Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991


```

;Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
;Reference number: I58280; MUID:91187610; PMID:1901405
;Accession: S22293
;Status: nucleic acid sequence not shown
;Molecule type: mRNA
;Residues: 1-670 <NT>
;Cross-references: EMBL:X54250; NID:G57519; PIDN:CAA38151.1; PID:G57520
;Note: The authors did not translate the codon for residue 1
;Superfamily: HIV-EP2 enhancer-binding protein
;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

y 3 VVPXGMSYS 11
||| |||
b 376 VVPAGLTYS 384

RESULT 9 82691
opoisomerase IV subunit XPI353 [imported] - Xylella fastidiosa (strain 9a5c)
;Species: Xylella fastidiosa
;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
;Accession: H82691
;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
ature 406, 151-157, 2000
;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
;Reference number: A82515; MUID:20365717; PMID:10910347
;Note: for a complete list of authors see reference number A59328 below
;Accession: H82691
;Status: preliminary
;Molecule type: DNA
;Residues: 1-749 <SIM>
;Cross-references: GB:AE003967; GB:AE003849; NID:G9106347; PIDN:AAF84162.1; GSPDB:GN001
;Experimental source: strain 9a5c
;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-
riones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
s-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fromm
D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
hado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
odrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
i.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
;Reference number: A59328
;Contents: annotation
;Gene: XPI353
;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase (
Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

y 2 EVVPXGMSY 10
||| |||
b 526 EVDFSGMSY 534

RESULT 10
34203
NA-binding protein PRDII-BP1 - human
;Alternate names: major histocompatibility complex enhancer-binding protein 1
;Species: Homo sapiens (man)
;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
;Accession: A34203; A34779
;Fan, C.M.; Maniatis, T.
enes Dev. 4, 29-42, 1990

```

```

;Title: A DNA-binding protein containing two widely separated zinc finger motifs that i
;Reference number: A34203; MUID:90169514; PMID:2106471
;Accession: A34203
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-2717 <FAN>
;Cross-references: EMBL:X51435; NID:G38017; PIDN:CAA35798.1; PID:G38018
R;Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
;Title: A large protein containing zinc finger domains binds to related sequence elemen
;Reference number: A34779; MUID:90205817; PMID:2108316
;Accession: A34779
;Status: preliminary; nucleic acid sequence not shown
;Molecule type: mRNA
;Residues: 501-1072; N' 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-16
;Cross-references: GB:M32019
;Superfamily: HIV-EP2 enhancer-binding protein
;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| |||
DB 2405 VVPAGLTYS 2413

RESULT 11
S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
;Alternate names: hypothetical protein O2612; hypothetical protein YOR303.3
;Species: Saccharomyces cerevisiae
;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
;Accession: S54619; S66879
R;de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
;Reference number: S54617
;Accession: S54619
;Molecule type: DNA
;Residues: 1-156 <DEH>
;Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123
R;de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
;Reference number: S66877
;Accession: S66879
;Molecule type: DNA
;Residues: 1-156 <DEW>
;Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA99201.1; PID:G1420111; MIPS:YOR01
;Experimental source: strain S288C
;Genetics:
;Cross-references: SGD:S0005539
;Map position: 15R
;Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
||| |||
DB 50 EVWFLGMDY 58

RESULT 12
H69491
cell division inhibitor (minD-2) homolog - Archaeoglobus fulgidus
;Species: Archaeoglobus fulgidus
;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
;Accession: H69491
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
;Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
;Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

```

Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S. Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Pyrococcus furiosus
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AE000970; GB:AE000782; NID:92689293; PIDN:AA89318.1; PID:G264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 9
 |||:||||
 Db 81 EVIPAGMS 88

RESULT 13

C82900
 Probable ABC substrate-binding protein, iron U359 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to Genbank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:96899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCsbp-5; U359
 A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
 |||||:||
 Db 135 BEVPHYLSY 144

RESULT 14

I40758
 hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.

J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohydrolase
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:Z36940; NID:9535805; PIDN:CAA95392.1; PID:9535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 ::|||
 Db 26 DIPPGMSY 34

RESULT 15

E90544
 50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
 A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: E90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:gl4089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPU 2610
 A:Genetic code: SGC3
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 |||:||||
 Db 68 VRPLGMSYS 76

Search completed: June 3, 2004, 12:00:03
 Job time : 9 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 3, 2004, 11:32:06 ; Search time 4.86667 seconds
(without alignments)
117.693 Million cell updates/sec

Title: US-09-909-164-47
Sequence: 52
Reference: 1 EVVXPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match %	Length	DB ID	Description
1	38	73.1	1058	1 CARB_FUSNN	Q8rg86 fusobacteri
2	37	71.2	877	1 SULH_SCHPO	Q74377 schizosacch
3	36	69.2	102	1 Y11K_TYDVA	P31619 tobacco yel
4	36	69.2	1498	1 Y1A9_CLOAB	Q04351 clostridium
5	35	67.3	2717	1 ZEP1_HUMAN	P15822 homo sapien
6	34	65.4	788	1 CY14_NEUCR	P23622 neurospora
7	34	65.4	1499	1 A10A_HUMAN	Q60312 homo sapien
8	33	63.5	116	1 RL20_MYCPU	Q98300 mycoplasma
9	33	63.5	165	1 YU49_ARCFU	Q28330 archaeglob
10	33	63.5	253	1 Y990_CAMJE	P45489 campylobact
11	33	63.5	280	1 CTX3_MOUSE	Q9d387 mus musculu
12	33	63.5	426	1 AROA_VIBCH	Q9kbr0 vibrio chol
13	33	63.5	466	1 CC37_SCHPO	Q94740 schizosacch
14	33	63.5	478	1 GSR2_HUMAN	Q9nzm5 homo sapien
15	33	63.5	890	1 BCN1_CLOPE	P08696 clostridium
16	33	63.5	1389	1 LTBS_MOUSE	Q8cg18 mus musculu
17	33	63.5	1394	1 LTBS_HUMAN	P22064 homo sapien
18	33	63.5	1401	1 RPOC_VIBCH	Q9kv29 vibrio chol
19	33	63.5	1595	1 LTBL_HUMAN	Q14766 homo sapien
20	33	63.5	1712	1 LTBL_RAT	Q00918 rattus norv
21	33	63.5	1713	1 LTBL_MOUSE	Q8cg19 mus musculu
22	32	61.5	288	1 CGD2_RAT	Q04827 rattus norv
23	32	61.5	289	1 CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	1 CGD2_MOUSE	P30280 mus musculu
25	32	61.5	291	1 CGD1_BRARE	Q90459 brachydanio
26	32	61.5	291	1 CGD1_XENLA	P50755 xenopus lae
27	32	61.5	291	1 CGD2_CHICK	P49706 gallus gall
28	32	61.5	291	1 CGD2_XENLA	P53782 xenopus lae
29	32	61.5	292	1 CGD1_CHICK	P55169 gallus gall
30	32	61.5	292	1 CGD3_HUMAN	P30281 homo sapien
31	32	61.5	295	1 CGD1_HUMAN	P24385 homo sapien
32	32	61.5	295	1 CGD1_MOUSE	P25322 mus musculu
33	32	61.5	295	1 CGD1_RAT	P39948 rattus norv

ALIGNMENTS

RESULT 1

ID	CARB_FUSNN	STANDARD;	PRT;	1058 AA.
AC	Q8RG86;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
GN	CARB OR FN0422.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;			
OC	Fusobacterium.			
OX	NCBI_TaxID=76856;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 25586;			
RX	MEDLINE=21886394; PubMed=11889109;			
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fongstein M., Kyripides N., Overbeek R.;			
RA	Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."			
RL	J. Bacteriol. 184:2005-2018(2002).			
CC	-I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.			
CC	-I- Cofactor: Binds 3 manganese ions per subunit (By similarity).			
CC	-I- PATHWAY: Arginine biosynthesis.			
CC	-I- SUBUNIT: Pyrimidine biosynthesis; first step.			
CC	-I- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).			
CC	-I- SIMILARITY: Belongs to the carb family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AE010554; AAL94625.1; ALT_INIT.			
DR	HAWAP; MF_01210; -; 1.			
DR	InterPro; IPR006275; CarA_L_glu.			
DR	InterPro; IPR005483; Cphase_L.			
DR	InterPro; IPR005479; Cphase_L_D2.			
DR	InterPro; IPR005480; Cphase_L_D3.			
DR	InterPro; IPR005481; Cphase_L_N.			
DR	InterPro; IPR004382; WGS_like.			
DR	Pfam; PF00289; Cphase_L_chain; 2.			
DR	Pfam; PF02786; Cphase_L_D2; 2.			
DR	Pfam; PF02787; Cphase_L_D3; 1.			

34	32	61.5	427	1	TOLB_HAEIN	P44677 haemophilus
35	32	61.5	529	1	ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1	S216_HUMAN	Q9y616 homo sapien
37	32	61.5	719	1	GSP_CRIFA	P90518 crithidia f
38	32	61.5	726	1	PRTF_HSV6U	P52384 human herpe
39	32	61.5	728	1	PRTF_HSV6Z	P52544 human herpe
40	32	61.5	759	1	PRTP_HSV6Z	P32784 saccharomyc
41	32	61.5	920	1	EDD_RAT	Q62671 rattus norv
42	32	61.5	993	1	VIA_TAV	P28931 tomato aspe
43	32	61.5	1377	1	RHSA_ECOLI	P16916 escherichia
44	32	61.5	1397	1	RHSC_ECOLI	P16918 escherichia
45	32	61.5	1411	1	RHSE_ECOLI	P16917 escherichia

```

DR Pfam: PF02142; MGS; 1.
DR PRINTS: PRO0098; CPSASE.
DR TIGRFAMs: TIGR01359; CPSASE1.lrg; 1.
DR PROSITE: PS00866; CPSASE_1; 2.
DR PROSITE: PS00867; CPSASE_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOXYMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP BIND 153 210 ATP (POTENTIAL).
FT NP BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXGMSYS 11
Db 190 EIVPENGLYNS 199
|:|:|:|:|
|:|:|:|:|

RESULT 2
SULH_SCHPO STANDARD; PRT; 877 AA.
AC 074377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable sulfate permease C3H7.02.
GN SPBC3H7.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hanlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Rivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vantreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Calibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

"The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002).

CC FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY

CC SIMILARITY).

CC SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53)

CC family.

CC SIMILARITY: Contains 1 STAS domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC EMBL: AL031261; CAA20298.1; -.

DR PIR: T40413; T40413.

DR GenBank: SPombe; SPBC3H7.02; -.

DR InterPro: IPR002645; STAS.

DR InterPro: IPR001902; Sulph. transpt.

DR Pfam: PF01740; STAS; 1.

DR Pfam: PF00916; Sulfate transp; 1.

DR TIGRFAMs: TIGR00815; sulp; 1.

DR PROSITE: PS01130; SLC26A; 1.

DR PROSITE: PS50801; STAS; 1.

KW Transport; Transmembrane.

FT TRANSMEM 133 153 POTENTIAL.

FT TRANSMEM 161 181 POTENTIAL.

FT TRANSMEM 186 206 POTENTIAL.

FT TRANSMEM 221 241 POTENTIAL.

FT TRANSMEM 243 263 POTENTIAL.

FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 329 349 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 424 444 POTENTIAL.

FT TRANSMEM 461 481 POTENTIAL.

FT TRANSMEM 484 504 POTENTIAL.

FT TRANSMEM 518 538 POTENTIAL.

FT TRANSMEM 543 563 POTENTIAL.

FT DOMAIN 594 747 STAS.

SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;

Best Local Similarity 77.8%; Pred. No. 9.4;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

Db 148 VVPQMSVA 156

|||:|:|:|

|||:|:|:|

RESULT 3

Y1LK_TYDVA STANDARD; PRT; 102 AA.

ID Y1LK_TYDVA

AC P31619; 1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-OCT-1993 (Rel. 27, Last annotation update)

DE Hypothetical 11.2 kDa protein.

DE VI.

OS Tobacco yellow dwarf virus (strain Australia) (TYDV).

OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.

OC NCBI_TaxID=31599;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92188538; PubMed=1546458;

RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;

RT "The nucleotide sequence of the infectious cloned DNA component of

RT tobacco yellow dwarf virus reveals features of geminiviruses

RT infecting monocotyledonous plants.";

RL Virology 187:633-642(1992).

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X51435; CAA35798.1; -
 DR PIR; A34203; A34203.
 DR PDB; 3ZNF; 15-JAN-92.
 DR PDB; 4ZNF; 15-JAN-92.
 DR PDB; 1BBO; 31-OCT-93.
 DR TRANSFAC; T00497; -
 DR Genew; HGNC:4920; HIVEPI.
 DR MIM; 194540; -
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF00096; zf-C2H2; 5.
 DR SMART; SM00355; ZnF_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear Protein; Repeat; 3D-structure.
 FT ZN_FING 406 428
 FT ZN_FING 434 456
 FT ZN_FING 958 981
 FT ZN_FING 2087 2109
 FT ZN_FING 2115 2139
 FT ZN_FING 2139 2159
 FT DOMAIN 803 806
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2095 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGXGMSYS 11
 DB 2405 VVPGAGTIS 2413

RESULT 6
 ID CY14_NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate permease II.
 GN CYS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OC NCBI_TaxID=5141;
 OX [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=91129256; PubMed=1825178;
 RA Ketter J.S., Jarai G., Fu Y.-H., Marzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of cys-14, the structural gene for sulfate permease II in
 RT Neurospora crassa.";
 RL Biochemistry 30:1780-1787 (1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RP MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate

permease II and a putative human tumour suppressor.";
 Trends Biochem. Sci. 19:19-19 (1994).
 CC FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC SUBCELLULAR LOCATION: Integral membrane protein.
 CC INDUCTION: Highly expressed, but only in cells subject to sulfur
 CC limitation, and it is turned on by the positive-acting Cys-3
 CC sulfur regulatory protein.
 CC MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC SIMILARITY: Belongs to the SLC26A/Sulp transporter (TC 2.A.53)
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; M59167; AAA33615.1; ALT_SEQ.
 PIR; A37956; A37956.
 DR InterPro; IPR001902; Sulph transpt.
 DR Pfam; PF00916; Sulfate transp; 1.
 DR TIGRFAMs; TIGR00815; sulp; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 71 91
 FT TRANSMEM 103 123
 FT TRANSMEM 128 148
 FT TRANSMEM 171 191
 FT TRANSMEM 193 213
 FT TRANSMEM 271 291
 FT TRANSMEM 326 346
 FT TRANSMEM 363 383
 FT TRANSMEM 451 471
 FT TRANSMEM 474 494
 FT CARBOHYD 23 23
 FT CARBOHYD 578 578
 SQ SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGXGMSYS 11
 DB 90 VVPGQAYYA 98

RESULT 7
 ID A10A_HUMAN STANDARD; PRT; 1499 AA.
 AC O60112; O96914;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1) (ATPVA)
 DE (Aminophospholipid-transporting ATPase VA).
 GN ATP10A OR ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 OX [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=21225279; PubMed=11326269;
 RA Meguro M., Kashiwagi A., Mitsuura K., Nakao M., Kondo I., Saitoh S.,
 RA Oshimura M.;
 RT "A novel maternally expressed gene, ATP10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome.";
 RL Nat. Genet. 28:19-20 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.

XX MEDLINE=21313119; PubMed=11353404;
 YA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
 YT "The human aminophospholipid-transporting ATPase gene ATP10C maps
 adjacent to UBE3A and exhibits similar imprinting expression.";
 ZL Am. J. Hum. Genet. 68:1501-1505(2001).
 [3]
 ZN SEQUENCE FROM N.A.
 ZP TISSUE=Skin;
 ZC MEDLINE=22389257; PubMed=12477932;
 ZQ Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 ZA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 ZB Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 ZC Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 ZD Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 ZE Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 ZF Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
 ZG Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 ZH Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 ZI Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 ZJ Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 ZK Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 ZL Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 ZM Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,
 ZN Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 ZO Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.,
 ZP Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 ZQ "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 ZR Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 ZS SEQUENCE OF 337-1499 FROM N.A.
 ZT TISSUE=Brain;
 ZU MEDLINE=98290545; PubMed=9628581;
 ZV Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 ZW Nomura N., Ohara O.;
 ZX "Prediction of the coding sequences of unidentified human genes. IX.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 ZY DNA Res. 5:331-39(1998).
 [5]
 ZAA CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 ZAB SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 ZAC TISSUE SPECIFICITY: Widely expressed, with highest levels in
 kidney, followed by lung, brain, prostate, testis, ovary and
 small intestine.
 ZAD DISEASE: Defects in ATP10A are a cause of Angelman syndrome (AS)
 (MM:105830); also known as 'happy puppet syndrome'. AS is
 characterized by features of severe motor and intellectual
 retardation, microcephaly, ataxia, frequent jerky limb movements
 and flapping of the arms and hands, hypotonia, hyperactivity,
 hypopigmentation, seizures, absence of speech, frequent smiling
 and episodes of paroxysmal laughter, and an unusual facies
 characterized by macrostomia, a large mandible and open-mouthed
 expression, a great propensity for protruding the tongue ('tongue
 thrusting'), and an occipital groove.
 ZAE SIMILARITY: Belongs to the cation transport ATPases family (P-type
 ATPases). Subfamily IV.
 [6]
 ZAF This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 or send an email to license@ebi.ac.uk).

DR EMBL; AY029493; AAK33100.1; JOINED.
 DR EMBL; AY029494; AAK33100.1; JOINED.
 DR EMBL; AY029495; AAK33100.1; JOINED.
 DR EMBL; AY029496; AAK33100.1; JOINED.
 DR EMBL; AY029497; AAK33100.1; JOINED.
 DR EMBL; AY029498; AAK33100.1; JOINED.
 DR EMBL; AY029499; AAK33100.1; JOINED.
 DR EMBL; AY029500; AAK33100.1; JOINED.
 DR EMBL; AY029501; AAK33100.1; JOINED.
 DR EMBL; AY029502; AAK33100.1; JOINED.
 DR EMBL; AY029503; AAK33100.1; JOINED.
 DR EMBL; BC052251; AAK33100.1; JOINED.
 DR EMBL; AB011138; BAA25492.1; -.
 DR EMBL; HGNC:13542; ATP10A.
 DR MIN; 605855; -.
 DR MIN; 105830; -.
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0004012; P:phospholipid-transporting ATPase activity; NAS.
 DR GO; GO:0008360; P:regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase_E1-E2.
 DR InterPro; IPR006539; Flippase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; CATALYTIC_PASE.
 DR PRINTS; PR00119; CATALYTIC_PASE.
 DR TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type; 6.
 DR PROSITE; PS00154; ATPase_E1-E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Multigene family.
 FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 87 106 POTENTIAL.
 FT DOMAIN 107 110 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 111 128 POTENTIAL.
 FT DOMAIN 129 309 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 310 332 POTENTIAL.
 FT DOMAIN 333 362 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 363 384 POTENTIAL.
 FT DOMAIN 385 1087 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1088 1108 POTENTIAL.
 FT DOMAIN 1109 1119 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1120 1140 POTENTIAL.
 FT DOMAIN 1141 1170 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1171 1192 POTENTIAL.
 FT DOMAIN 1193 1199 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1200 1222 POTENTIAL.
 FT DOMAIN 1223 1228 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1229 1249 POTENTIAL.
 FT DOMAIN 1250 1267 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1268 1292 POTENTIAL.
 FT DOMAIN 1293 1499 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 427 427 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 1031 1031 MAGNESIUM (BY SIMILARITY).
 FT METAL 1035 1035 MAGNESIUM (BY SIMILARITY).
 FT DOMAIN 467 470 POLY-GLU.
 FT CONFLICT 388 388 Q -> R (IN REF. 4).
 SQ SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
 Query Match 65.48; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred.No. 70;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 |||||
 DB 469 BEVVPXGMSYS 479
 RESULT 8
 RL20 MYCPU
 ID RL20 MYCPU STANDARD; PRT; 116 AA.
 AC Q980V0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

DE 50S ribosomal protein L20.
GN RPLT OR MYPU 2610.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=2167165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153(2001).
CC -!- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
CC necessary for the in vitro assembly process of the 50s ribosomal
CC subunit. It is not involved in the protein synthesizing functions
CC of that subunit (by similarity).
CC -!- SIMILARITY: Belongs to the L20P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445563; CA13434.1; -.
DR FJIR; E90544; E90544.
DR MyPulList; MYPU 2610; -.
DR HAMAP; MF_00382; -.
DR InterPro; IPR005813; Ribosomal L20.
DR InterPro; IPR005812; Ribosomal L20b/o.
DR Pfam; PF00453; Ribosomal L20; 1.
DR PRINTS; PR00062; RIBOSOMALL20.
DR ProDom; PD002389; L20; 1.
DR TIGRfams; TIGR01032; rplT_bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 8.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 68 VRPLGMSYS 76

RESULT 9
Y049_ARCFU ARCFU STANDARD; PRT; 165 AA.
AC 028330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glöck A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000968; AAB89307.1; -.
DR FJIR; D69493; D69493.
DR TIGR; AF1949; -.
DR KW Hypothetical protein; Transmembrane; Complete proteome.
DR TRANSMEM 7 27 POTENTIAL.
DR TRANSMEM 141 161 POTENTIAL.
DR SEQUENCE 165 AA; 17588 MW; BBC17054810ADB8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 60 RESIPDGASY 69

RESULT 10
Y990_CAMJE CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CJ0990C.
GN CJ0990C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rejandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whithead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences";
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli.";
RL J. Bacteriol. 177:2396-2402(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL139076; CAB73246.1; -
EMBL; Z36940; CAA85392.1; -
PIR; C81374; C81374.
PIR; I40758; I40758.
HYPOTHETICAL PROTEIN; Complete proteome.
SEQUENCE 253 AA; 29783 MW; F96DJFF326F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

2 EVVPGMSY 10

185 DFFSGMSY 193

RESULT 11

TX3_MOUSE

ID CTX3_MOUSE STANDARD; PRT; 280 AA.

Q9D387; Q9CX04;

28-FEB-2003 (Rel. 41, Created)

10-OCT-2003 (Rel. 42, Last annotation update)

Protein C20orf103 homolog precursor.

C20orf103.

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI TaxID=10090;

SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

STRAIN=C57BL/6J; TISSUE=Embryonic head;

MEDLINE=21085660; PubMed=1127851;

Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,

Sato T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,

Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

Kucler P., Lewis S., Matsuo Y., Nikaido I., Peole G., Quackenbush J.,

Schriml L.M., Scahill P., Suzuki R., Tomita M., Wagner L., Washio T.,

Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

Suzuki H., Toyooka K., Wang K., Weitz C., Whittaker C., Wilming L.,

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,

Hayashizaki Y.,

"functional annotation of a full-length mouse cDNA collection."

Nature 409:685-690(2001).

SEQUENCE FROM N.A. (ISOFORM 2).

TISSUE=Mammary fibroblast;

MEDLINE=22388257; PubMed=12477932;

Klausner R.L., Feingold B.A., Grouse L.H., Derge J.G.,

Strausberg R.L., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,

Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Xie H.,

Diachenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M.J., Udell T.B., Toshitsugu S., Carninci P., Prange C.,

Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Pahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U.A., Smallos D.E.,
Scherer A., Schein J.E., Jones S.J.M., Marra M.A.,
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
-1- ALTERNATIVE PRODUCTS:
Svent-Alternative splicing; Named isoforms=2;
Names=1;
IsoId=Q9D387-1; Sequence=Displayed;
Names=2;
IsoId=Q9D387-2; Sequence=VSP_003820;
-1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
in positions 174 and 239.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AK014127; BAB29169.1; -
EMBL; AK018222; BAB31124.1; ALT_FRAME.
EMBL; BC004791; AAH04791.1; -
MGD; MGI:1920368; 3110035N03Rik.
MGD; MGI:1923411; 6330527006Rik.

Transmembrane; Signal; Alternative splicing.

SIGNAL 1 29 POTENTIAL.

CHAIN 30 280 PROTEIN C20orf103 HOMOLOG.

DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).

TRANSMEM 236 256 POTENTIAL.

DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).

CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).

VARSPIC 1 118 Missing (in isoform 2).

CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).

CONFLICT 230 230 Q -> P (IN REF. 1; BAB31124).

CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).

SEQUENCE 280 AA; 31721 MW; FALID7BF9FD5CCEP CRC64;

Query Match 63.5%; Score 33; DB 1; Length 280;

Best Local Similarity 75.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

3 VVPXGMSY 10

173 VTPAGMSY 180

RESULT 12

AROA_VIBCH

ID AROA_VIBCH STANDARD; PRT; 426 AA.

AC Q9XRB0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 3-phosphohexikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).

AROA OR VC1732.

GN AROA OR VC1732.

OS AROA OR VC1732.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OC NCBI_TaxID=666;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=El Tor N16961 / Serotype O1;
RX  MEDLINE=20406833; PubMed=10952301;
RA  Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA  Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA  Gill S.R., Nelson K.E., Read T.D., Tettelin H., Dragoi I., Sellers P.,
RA  Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA  McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA  Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA  Fraser C.M.;
RA  "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT  cholerae.";
RL  Nature 406:477-483 (2000).
CC  -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC  phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC  -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC  sixth step.
CC  -!- SUBUNIT: Monomer (By similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -!- SIMILARITY: Belongs to the EPSP synthase family.
CC  This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AB004251; AAF94882.1; -.
CC  PIR; D82163; D82163.
CC  TIGR; VCL1732; -.
CC  HAMAP; MF_00210; -.
CC  InterPro; IPR006264; AroA.
CC  InterPro; IPR001986; EPSP synth.
CC  Pfam; PF00275; EPSP synthase; 1.
CC  ProDom; PD001867; EPSP synthase; 1.
CC  TIGRfam; TIGR01356; aroA; 1.
CC  PROSITE; PS00104; EPSP SYNTHASE 1; 1.
CC  PROSITE; PS00885; EPSP SYNTHASE 2; 1.
CC  Aromatic amino acid biosynthesis; Transferase; Complete proteome.
KW  AROMATIC AMINO ACID BIOSYNTHESIS; TRANSFERASE; COMPLETE PROTEOME.
SQ  SEQUENCE 426 AA; 46101 MW; 38952D6463BFE1C3 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
DB 223 EFVIPAGQSY 232

RESULT 13
CC37_SCHPO STANDARD; PRT; 466 AA.
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE subunit) (Cell division control protein 37).
GN CDC37 OR SPBC986.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Westwood P.K., Preston N.C., Fantes P.A.;
RA "Schizosaccharomycetes pombe cdc37 gene.";
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=972;
RX  MEDLINE=21848401; PubMed=11859360;
RA  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA  Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA  Holtroyd S., Hornsby T., Howarth S., Huckle B.J., Hunt S., Jagels K.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA  Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA  Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA  Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA  Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA  Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerruti L., Lowe T., McComb W.R., Paulsen I., Potashkin J.,
RA  Szpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT  The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 415:871-880 (2002).
RN  [3]
RP  FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX  MEDLINE=22745456; PubMed=12861001;
RA  Tatebe H., Shiozaki K.;
RT  Identification of cdc37 as a novel regulator of the stress-responsive
RT  mitogen-activated protein kinase.";
RL  Mol. Cell. Biol. 23:5132-5142 (2003).
CC  -!- FUNCTION: Co-chaperone that binds to numerous kinases and promotes
CC  their interaction and promotion of their activity.
CC  -!- SUBUNIT: Forms a complex with Hsp90. Interacts with sty1.
CC  -!- SUBCELLULAR LOCATION: Nuclear, and cytoplasmic. When in the
CC  nucleus associated with chromatin.
CC  -!- SIMILARITY: Belongs to the CDC37 family.
CC  This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AJ132377; CAB38758.1; -.
CC  EMBL; AJ132376; CAB38757.1; -.
CC  EMBL; AL049769; CAB42371.2; -.
CC  PIR; T43653; T43653.
CC  GenBank; SPombe; SPBC986.10; -.
CC  InterPro; IPR004318; Cdc37.
CC  Pfam; PF03234; Cdc37; 1.
KW  Chaperone; Cell division; Cell cycle; Nuclear protein.
SQ  SEQUENCE 466 AA; 52554 MW; 647238B34CABB3C5 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 466;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVVPXGMSY 10
DB 98 DSAIPGMSY 107

RESULT 14

```

IS22 HUMAN STANDARD; PRT: 478 AA.
D -GSR2 HUMAN STANDARD; PRT: 478 AA.
C QNZM5; Q9B7C6; Q9HAX6; Q9NP14; Q9NP12;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Glioma tumor suppressor candidate region gene 2 protein (p60).
N GTSR2.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
N [1]
P SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
X MEDLINE=20175430; PubMed=10708517;
A Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
A Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
A Scheithauer B.W., Louis D.N., Jenkins R.B.;
T "A transcript map of the chromosome 19q-arm glioma tumor suppressor
T region.";
L Genomics 64:44-50 (2000).
N [2]
P SEQUENCE FROM N.A.
C TISSUE=Muscle;
X MEDLINE=22398257; PubMed=12477932;
A Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
A Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
A Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
A Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,
A Diatchenko L., Marusik K., Farmer A., Rubin G.M., Hong L.,
A Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
A Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
A Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
A Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
A Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hukly S.W.,
A Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
A Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
A Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
A Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
A Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
A Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
A Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
T "Generation and initial analysis of more than 15,000 full-length
T human and mouse cDNA sequences.";
L Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
N [3]
P SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
X MEDLINE=99214318; PubMed=10196275;
A Bruni R., Fineschi B., Ogle W.O., Roizman B.;
T "A novel cellular protein, p60, interacting with both herpes simplex
T virus 1 regulatory proteins ICP22 and ICP0 is modified in a
T cell-type-specific manner and is recruited to the nucleus after
T infection.";
J Virol. 73:3810-3817 (1999).
N [4]
P SEQUENCE OF 12-478 FROM N.A.
A Andreu N., Estivill X., Escarceller M., Sumoy L.;
L Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
N [5]
P SEQUENCE OF 218-477 FROM N.A.
C TISSUE=Testis;
X Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
L Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
C -!- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
C -!- SUBCELLULAR LOCATION: Nuclear.
C -!- TISSUE SPECIFICITY: Expressed at high levels in heart and
C pancreas, moderate levels in placenta, liver, skeletal muscle, and
C kidney, and low levels in brain and lung.
C -!- SIMILARITY: Belongs to the GTSR2 family.

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@sib-sib.ch).

CC ENBL; AF182076; AAF62873.1; -;
CC ENBL; BC004229; AAH04229.1; -;
CC ENBL; BC006311; AAH06311.1; -;
CC ENBL; BC010095; AAH10095.1; -;
CC ENBL; AF296124; AAG30413.1; -;
CC ENBL; AL359335; CAB94786.1; -;
CC ENBL; AL359336; CAB94787.1; -;
CC ENBL; AL122063; CAB59242.1; -;
CC SWISS-2DPAGE; Q9NZM5; HUMAN.
CC Genew; HGNC:4933; GLTSCR2.
CC MIM; 605691; -;
CC GO; GO:0005622; C:intracellular; NAS.
CC Nuclear protein; Polymorphism.
CC VARIANT 389 389 R -> Q.
CC CONFLICT 4 6 /FTID=VAR_011496.
CC CONFLICT 9 9 GGS -> HEG (IN REF. 2; AAH04229).
CC CONFLICT 146 191 R -> R (IN REF. 3).
CC CONFLICT 198 215 RKEQLWEKLAKQGLPREVRAQARLLNPATRAKPGPQD
CC CONFLICT 235 235 TVERP -> SGRSYGRSWPSPGASPGGAQGFSPVAQPCFN
CC CONFLICT 417 417 KGNPAPGHR1AA (IN REF. 3).
CC CONFLICT 433 477 SDNPLDRPLVGQDEFFLE -> LNNPDKPVVWPGCLPFG
CC CONFLICT 434 478 (IN REF. 3).
CC CONFLICT 478 478 A -> S (IN REF. 2; AAH04229).
CC CONFLICT 478 478 D -> H (IN REF. 3).
CC CONFLICT 478 478 PEGNLRDRFKSFRQNMIEPRERAKFRKVKVLEKRAF
CC CONFLICT 478 478 REIQ -> VLVTSRCGAPCPVMTSLPLVPPRGYRHHGCP
CC CONFLICT 478 478 WAGPVGMPRG (IN REF. 5).
CC CONFLICT 478 478 EGNILDRPK3FORNMIEPRERAKFRKVKVLEKRAF
CC CONFLICT 478 478 ETQL -> RGHSFETGSRFEGGI (IN REF. 3).
CC SEQUENCE 478 AA; 54417 MW; 7F18923E348CB52B CRC64;
Query Match 63.5%; Score 33; DB 1; Length 478;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVPGKMSYS 11
Db 239 EVAPAGASYN 248
RESULT 15
ID -BCN5 CLOPE STANDARD; PRT; 890 AA.
AC P08656;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Bacteriocin BCN5.
GN BCN.
OS Clostridium perfringens.
OG Plasmid pIP404.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=88336297; PubMed=2901768;
RA Garner T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garner T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium

RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene.";
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
vivo and in vitro.";
RL Mol. Microbiol. 2:607-614(1988).
CC -!- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -!- INDUCTION: BY UV irradiation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; M14481; AAA98248.1; -;
DR ENBL; M32882; AAA98249.1; -;
DR PIR; A30481; A30481.
DR InterPro; IPR000834; Peptidase_M14.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_carboOept; 1.
DR SMART; SM00287; SH3b; 3.
DR Antibiolic; Bacteriocin; Plasmid.
KW DOMAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4E5E971C31C6C6 CRC64;
SQ

Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 170 EVVPGGFTY 178

Search completed: June 3, 2004, 11:49:56
Job time : 4.86667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2004, 11:35:06 ; Search time 29.8667 Seconds
(without alignments)
116.206 Million cell updates/sec

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 BEWVFXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_arChaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_proteid:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	75.0	1044	16	Q8DIH0	Q8diho synchococ
2	38	73.1	344	16	Q815A7	Q815a7 bacillus ce
3	38	73.1	387	16	Q98EX1	Q98fx1 rhizobium 1
4	38	73.1	3472	1	O74056	O74056 cnarchaeum
5	37	71.2	840	3	Q9URX8	Q9ury8 schizosacch
6	37	71.2	1123	16	Q8EW04	Q8ew04 mycoplasma
7	36	69.2	471	11	Q8R126	Q8r126 mus musculus
8	36	69.2	484	11	Q8VD18	Q8vdi8 mus musculus
9	36	69.2	484	11	Q8BTX4	Q8btx4 mus musculus
10	36	69.2	484	11	Q8BK35	Q8bk35 mus musculus
11	36	69.2	559	16	Q839T9	Q839t9 enterococcu
12	36	69.2	1399	16	Q889X7	Q889x7 pseudomonas
13	35	67.3	225	10	Q40129	Q40129 lycopersico
14	35	67.3	245	16	Q7V6Q4	Q7v6q4 prochloroco
15	35	67.3	425	5	Q9XVK4	Q9xvk4 caenorhabdi
16	35	67.3	495	11	Q8C1D7	Q8c1d7 mus musculus

Q7z6r0 homo sapien
Q9bh83 plasmodium
Q9bha5 plasmodium
Q8i5s7 plasmodium
Q01487 rattus ratt
Q8pmi6 xanthomonas
Q8pat2 xanthomonas
Q9pdm6 xyella fas
Q14122 homo sapien
Q12479 saccharomyc
Q971e2 sulfolobus
Q28342 archaeoglob
Q96mul homo sapien
Q92md6 rhizobium m
Q8iym3 homo sapien
Q98bp5 rhizobium 1
Q9pqr2 ureaplasma
Q7uwu7 rhodopirell
Q9urra penicillium
Q8g4i5 bifidobacte
Q8xt05 raietonia s
Q87kq5 vibrio para
Q8tx62 methanopyru
Q8ppp5 xanthomonas
Q8vua8 lactococcus
Q8ktq4 candidatus
Q8r5e6 fusobacteri
Q7sy67 xenopus lae
Q7u552 synchococc

ALIGNMENTS

RESULT 1

Q8DIH0 PRELIMINARY; PRT; 1044 AA.
ID Q8DIH0
AC Q8DIH0
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1";
RL DNA Res. 9:123-130(2002).
DR EMBL; AF005374; BAC09170.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001036; Acrlflin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_tran; i.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00915; 2A0602; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

Query Match 75.0%; Score 39; DB 16; Length 1044;
Best Local Similarity 53.6%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 BEVVPXGMSYS 11
DB 843 BEVLPNGIGIS 853

RESULT 2
Q815A7 PRELIMINARY; PRT; 344 AA.
AC Q815A7; 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter substrate-binding protein.
GN BC5259.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Golsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
RL EMBL; AE017015; AAP12123.1; -.
DR InterPro; IPR00437; Prok_lipoptot S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 344 AA; 38539 MW; C55268ACB7225995 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 344;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 10
DB 152 BEIAPGLSY 161

RESULT 3
Q88FX1 PRELIMINARY; PRT; 387 AA.
AC Q88FX1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hippurate hydrolase.
GN M183583.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
RL EMBL; AF003002; BAB50445.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase_M20.

DR Pfam; PF01546; Peptidase M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 10
DB 367 DEAIPIHMSY 376

RESULT 4
Q74056 PRELIMINARY; PRT; 3472 AA.
AC Q74056;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
OC Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B.
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL; AF083072; AAC62699.1; -.
DR PIR; T31308; T31308.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR001680; WD40.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
KW Hypothetical protein.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 2294 EDVIPRGISFS 2304

RESULT 5
Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC89.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972h-;
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
```

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

EMBL; AL132779; CAB60015.1; -;
PIR; T39116; T39116;
GeneDB SPombe; SPAC869.05C; -;
GO; GO:0016020; C:membrane; IEA.
GO; GO:0008271; F:sulfate porter activity; IEA.
GO; GO:0008272; P:sulfate transport; IEA.
InterPro; IPR002645; SPAS
InterPro; IPR001902; Sulph_transpt.
Pfam; PF01740; STAS; 1.
Pfam; PF00916; Sulfate_transp; 1.
TIGRFAMs; TIGR00815; sulp; 1.
PROSITE; PS50801; STAS; 1.
SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;

Best Local Similarity 77.8%; Pred.No. 59; Mismatches 1; Indels 0; Gaps 0;

2Y 3 VVPXGMSYS 11

135 VVPQMSYA 143

RESULT 6

Q8EWD4

PRELIMINARY; PRT; 1123 AA.

01-MAR-2003 (TrEMBLrel. 23, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
MYPE 2560 paralog, 57%.
MYPE2710.
Mycoplasma penetrans.
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=28227;
SEQUENCE FROM N.A.
STRAIN-HF-2;
MEDLINE=22354719; PubMed=1246655;
Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
"The complete genomic sequence of Mycoplasma penetrans, an
intracellular bacterial pathogen in humans."
Nucleic Acids Res. 30:5293-5300(2002).
EMBL; AP004171; BAC4062.1; -;
InterPro; IPR008985; ConA_like_lcg1.
InterPro; IPR007326; Lipoprotein_17.
Pfam; PF04200; Lipoprotein_17; 3.
Complete proteome.
SEQUENCE 1123 AA; 123636 MW; A4D707330E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;

Best Local Similarity 70.0%; Pred.No. 81; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11

658 EIVPGLSYS 667

RESULT 7

Q8R126

PRELIMINARY; PRT; 471 AA.

01-JUN-2002 (TrEMBLrel. 21, Created)
01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
Hypothetical protein (Fragment).
GLTSCR2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

SEQUENCE FROM N.A.
TISSUE=Liver;
Strausberg R.;
Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC025810; AAH25810.1; -;
MGI; MGI:2154441; GLTscr2.
Hypothetical protein.
NON TER 1
SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;

Best Local Similarity 60.0%; Pred.No. 51; Mismatches 6; Conservative 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11

226 EVIPAGASYN 235

RESULT 8

Q8VD18

PRELIMINARY; PRT; 484 AA.

01-MAR-2002 (TrEMBLrel. 20, Created)
01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Similar to glioma tumor suppressor candidate region gene 2.
GLTSCR2 OR AW53641.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
SEQUENCE FROM N.A.
TISSUE=Salivary gland;
Strausberg R.;
Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
EMBL; BC017637; AAH17637.1; -;
MGI; MGI:2154441; GLTscr2.
SEQUENCE 484 AA; 55835 MW; BBB45FB34BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;

Best Local Similarity 60.0%; Pred.No. 53; Mismatches 6; Conservative 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11

239 EVIPAGASYN 248

RESULT 9

Q8BTX4

PRELIMINARY; PRT; 484 AA.

01-MAR-2003 (TrEMBLrel. 23, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Similar to glioma tumor suppressor CANDIDATE region gene 2
protein.
GLTSCR2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

SEQUENCE FROM N.A.

STRAIN=NOB; TISSUE=Thymus;

MEDLINE=22354683; PubMed=12466851;

The FANTOM Consortium,

the RIKEN Genome Exploration Research Group Phase I & II Team;

"Analysis of the mouse transcriptome based on functional annotation of

```

RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002);
DR ENBL; AK088461; BAC40367.1; -.
DR MGD; MGI:2154441; Gliccr2.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5BECAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 10
Q8BK35 PRELIMINARY; PRT; 484 AA.
AC Q8BK35;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
DE protein.
DE GITSCK2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/60; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002);
DR ENBL; AK077341; BAC36760.1; -.
DR MGD; MGI:2154441; Gliccr2.
SQ SEQUENCE 484 AA; 55792 MW; EB67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 239 EVIPAGASYN 248

RESULT 11
Q839T9 PRELIMINARY; PRT; 559 AA.
AC Q839T9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phorbolone binding protein, putative.
DE EF0063.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V593 / ATCC 700802;
RX MEDLINE=22550857; PubMed=1263927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;

```

```

RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR ENBL; AE016947; AAO79943.1; -.
DR TIGR; EF0063; -.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok lipoprot_5.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 559 AA; 61476 MW; CCL5418D33D53DE7 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 559;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 351 LIPEGMSYS 359

RESULT 12
Q889X7 PRELIMINARY; PRT; 1399 AA.
AC Q889X7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA-directed RNA polymerase, beta' subunit.
DE RPOC OR PSPT00620.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Barry K., Utterback T., Van Aken S., Feldblyum T., Winn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.;
RA Submitted (MAR-2003) to the ENBL/GenBank/DBJ databases.
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the ENBL/GenBank/DBJ databases.
DR ENBL; AE016858; AAO54162.1; -.
DR TIGR; PSPT00620; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR DNA-directed RNA polymerase, Complete proteome.
KW Complete proteome.
SQ SEQUENCE 1399 AA; 154733 MW; 26178B653102ADE CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1399;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 10
Db 581 QVVPAGLSY 589

```


Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVFXGMSYS 11
Db 179 QVVPVGLGYS 188

RESULT 15

Q9XVK4 PRELIMINARY; PRT; 425 AA.
ID Q9XVK4
AC Q9XVK4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81109; CAB03241.1; -.
DR PIR; T24111; T24111.
DR WormPep; R10D12.10; CE12690.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004572; F:protein kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVFXGMSY 10
Db 335 EQIVPGGLQY 344

Search completed: June 3, 2004, 11:57:35
Job time : 29.8667 secs

Best Local Similarity 67.3%; Score 35; DB 10; Length 225;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVFXGMSYS 11
Db 32 DEVVPNGKYA 42

RESULT 14

Q7V6Q4 PRELIMINARY; PRT; 245 AA.
ID Q7V6Q4
AC Q7V6Q4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospholipid and glycerol acyltransferase (From 'motifs_6.msf').
EN PNT1092.
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.P., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb S.E., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation."
RL Nature 424:1042-1047(2003).
DR EMBL; BX572098; CAE21267.1; -.
KW Acyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 245 AA; 26907 MW; 106F7C4CBEC6427 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 245;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVFXGMSYS 11
Db 32 DEVVPNGKYA 42